The Dissertation Committee for Reza Soheilifard Certifies that this is the approved version of the following dissertation:

Mechanical Models of Proteins

Committee:

Gregory J. Rodin, Supervisor

Dmitrii E. Makarov, Co-Supervisor

Krishnaswamy Ravi-Chandar

Jeffrey K. Bennighof

Ron Elber

Mechanical Models of Proteins

by

Reza Soheilifard, B.E; M.E.

Dissertation

Presented to the Faculty of the GraduateSchool of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2011

Dedication

To my lovely wife Zahra

To my dear parents

To my beautiful children Amir Hossein and Zainab

and

To the memory of my brother-in-law Mohammad

Acknowledgements

I would never have been able to finish my dissertation without God's grace, the guidance of my supervisors, support of my wife, and help of my friends.

I am greatly thankful to my supervisors, Professor Gregory J. Rodin and Professor Dmitrii E. Makarov, for all their assistance, support and guidance throughout my research. Additionally, I would like to thank my committee members for their very helpful discussions, comments and suggestions.

My deepest gratitude is due to my beloved wife Zahra, for her endless love, help and support during all these years of my PhD study. She did an outstanding job, coming that far away from home, bringing up our kids, inspiring and encouraging me to finish my dissertation successfully.

I am also grateful to my dear parents for their love and encouragement, as they set my feet on the path that has brought me to this point in my life.

Last but not least, I would like to acknowledge the invaluable help and assistance of all my Iranian friends in Austin who, like a big family, supported us during my studies, and this research would not have been completed without them.

Reza Soheilifard

November 2011

Mechanical Models of Proteins

Reza Soheilifard, Ph.D.

The University of Texas at Austin, 2011

Supervisors: Gregory J. Rodin and Dmitrii E. Makarov

In general, this dissertation is concerned with modeling of mechanical behavior of protein molecules. In particular, we focus on coarse-grained models, which bridge the gap in time and length scale between the atomistic simulation and biological processes. The dissertation presents three independent studies involving such models. The first study is concerned with a rigorous coarse-graining method for dynamics of linear systems. In this method, as usual, the conformational space of the original atomistic system is divided into master and slave degrees of freedom. Under the assumption that the characteristic timescales of the masters are slower than those of the slaves, the method results in Langevin-type equations of motion governed by an effective potential of mean force. In addition, coarse-graining introduces hydrodynamic-like coupling among the masters as well as non-trivial inertial effects. Application of our method to the long-timescale part of the relaxation spectra of proteins shows that such dynamic coupling is essential for reproducing their relaxation rates and modes.

The second study is concerned with calibration of elastic network models based on the so-called B-factors, obtained from x-ray crystallographic measurements. We show that a proper calibration procedure must account for rigid-body motion and constraints imposed by the crystalline environment on the protein. These fundamental aspects of protein dynamics in crystals are often ignored in currently used elastic network models,

leading to potentially erroneous network parameters. We develop an elastic network model that properly takes rigid-body motion and crystalline constraints into account. This model reveals that B-factors are dominated by rigid-body motion rather than deformation, and therefore B-factors are poorly suited for identifying elastic properties of protein molecules. Furthermore, it turns out that B-factors for a benchmark set of three hundred and thirty protein molecules can be well approximated by assuming that the protein molecules are rigid.

The third study is concerned with the polymer mediated interaction between two planar surfaces. In particular, we consider the case where a thin polymer layer bridges two parallel plates. We consider two models of monodisperse and polydisperse for the polymer layer and obtain an analytical expression for the force-distance relationship of the two plates.

Table of Contents

List of Tables

List of Figures

Figure 4.1: Schematics of the formation of a tail, a bridge, a loop, and a train between two surfaces. 87

Chapter One
Introduction

Proteins are large macromolecules made of amino acids as their building blocks. Proteins are the most versatile macromolecules in living systems and they are crucial to all biological functions. Therefore there is a tremendous body of literature concerned with proteins and their behavior. Of particular interest to this study is protein dynamics as it plays a major role in many biological processes[1, 2].

Due to protein complexity, their dynamics is usually simulated using atomistic or molecular models. In those simulations, a minimum time step is typically 10^{-15} s, which is necessity to resolve atomistic vibrations whose period is about 10^{-14} s [3]. In contrast, most biological processes involve time scales in the range between 10^{-6} and 10^{3} s. This gap cannot be closed with currently available computing resources, which allow atomistic simulations in the range between 10^{-9} and 10^{-6} s [4, 5]. In general time scales accessible to atomistic simulations depend on the system size, which can be as large as hundreds of thousands of atoms for a single macromolecule.

One can close the length and time scale gaps between atomistic simulations and biological processes by considering simplified or coarse-grained models. Those models must offer an optimal trade-off between accuracy and efficiency, so that, on the one hand, the coarse-grained system is significantly smaller than the original system, and, on the other hand, the coarse-grained system retains essential features of the original system. Such models constitute the primary concern of this dissertation. In particular, we address

three topics: (i) rigorous coarse-graining procedures for linear systems, (ii) calibration of elastic network models, and (iii) collective behavior of random-walk chains in constrained volumes.

1.1. Protein structure

Due to their diversity, the structural classification of proteins is organized in a hierarchical manner[6]. The first level of this hierarchy of the *primary structure* identifies the sequence of the amino acids (or residue) forming the protein backbone (Fig. 1.1a). Each residue is formed by an amine group, an acidic carboxylic group, and a side-chain (R group). The carbon atoms belonging to the backbone are called the α–carbon atoms and the carbon atoms in the side chains next to α–carbons are called β-carbon atoms. In coarse-grained models, it is widely accepted to regard the α-carbons as the sole representatives of their residues [7-9].

Local substructures of amino acids, stabilized by hydrogen bonds, form the second level of the hierarchy of the *secondary structure*(Fig. 1.1b). Common secondary structures are α-helices, and pleated or β-sheets. On the third level of the hierarchy, one identifies the *tertiary structure*or the three-dimensional shape of the backbone governed by the secondary structures and cross linking among them. The tertiary structure is of critical importance from the mechanical point of view, as it governs the protein geometry and ultimately mechanical properties. Finally, the *quaternary structure*is a larger assembly of several protein or polypeptide chains, usually called subunits.

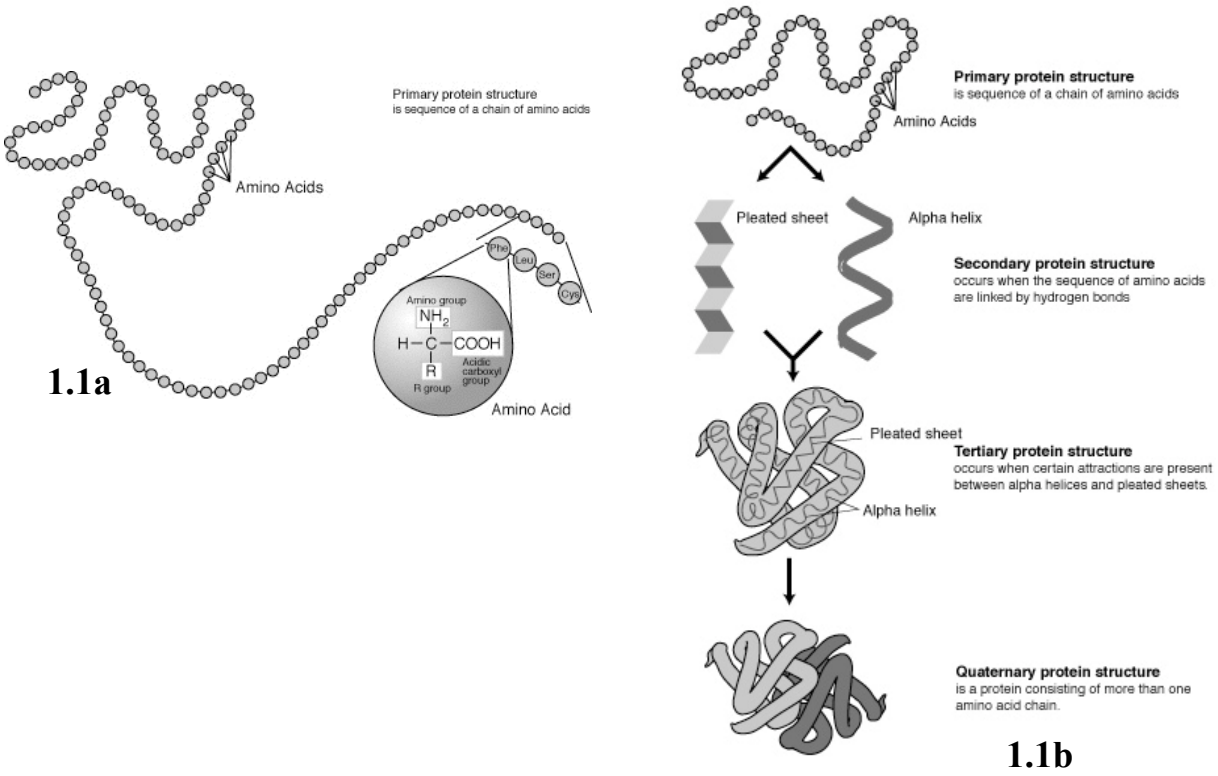


Figure 1.1 a) Schematic of the primary structure. b) Hierarchical protein structure.

[courtesy of Darryl Leja, National Human Genome Research Institute (NHGRI)]

The three-dimensional structure of a protein was first identified by a British biochemist John Kendrew [6] in 1960. Since then, the three-dimensional structure has been identified for tens of thousands of proteins. Today, these structures can be freely accessed over the internet, through the protein data bank (www.rcsb.org). The information about these structures has been determined using experimental techniques, like X-ray crystallography, nuclear magnetic resonance, and electron microscopy. For each structure in the data bank, the corresponding file contains Cartesian coordinates of the atoms. Some files also contain the information pertaining to the so-called B-factors, which is widely used for calibrations of elastic network models. This topic will be addressed in detail in Chapter 3.

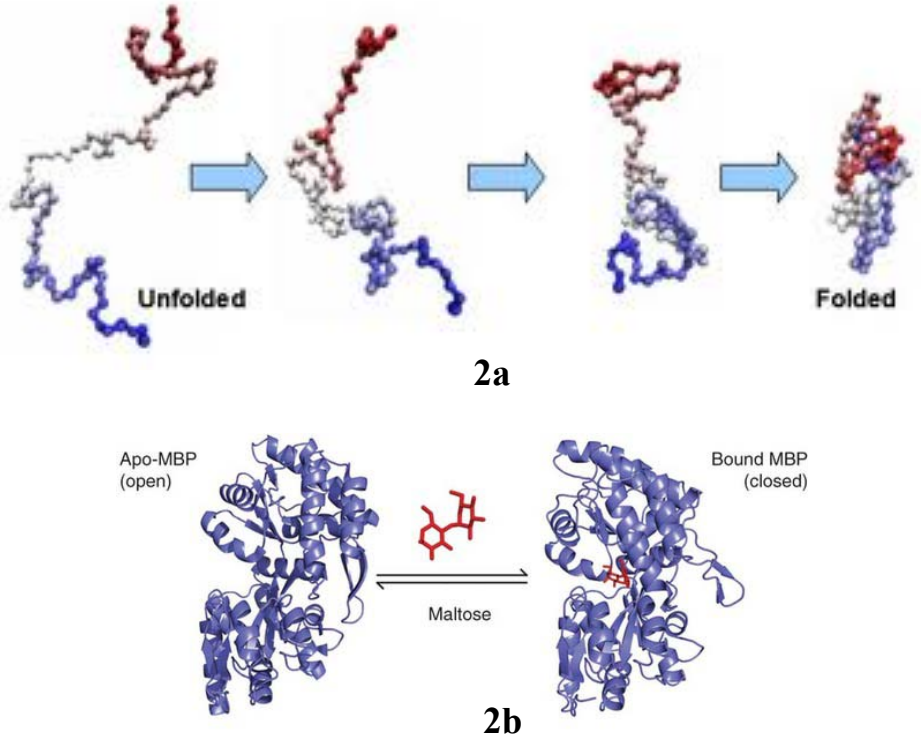


Figure 1.2. a) Schematic of the process of protein folding b) conformational change of a protein upon binding to a substrate.[modified from Ref.[10]]

1.2. Protein dynamics

It is widely accepted that the three-dimensional shape or *conformation* can change either spontaneously, due to thermal motion, or due to external factors, usually with environmental changes. The most important change is *folding*. In this process, a protein undergoes structural changes from a random coil into a unique three dimensional structure, called *native conformation*. It is in this folded form that the protein is functional. Figure 1.2a shows a schematic for protein folding. Folding and unfolding have been the subject of many studies [11-15]. Atomistic simulation of the entire folding process requires high-fidelity models and long simulation times [16, 17]. Other

conformational changes are associated with protein binding to a ligand [18, 19] (Fig. 1.2b) and protein propulsion [20]. Most of these conformational changes involve collective motion of protein subunits[20]. Hence, their dynamics should be well described using appropriate coarse-grained models, like those presented in Chapters 2 and 3.

1.3. Modeling proteins dynamics

1.3.1. Molecular Dynamics

Molecular dynamics is the natural choice for proteins simulations. The first molecular dynamics simulation was done by Alder and Wainwright in 1957[21] who studied a system of hard spheres. Later in 1977, McCammon and co-workers[22] performed the first molecular dynamics simulation of a protein, bovine pancreatic trypsin inhibitor (BPTI). The basic idea behind molecular dynamics is to numerically integrate equations of motion for a system of particles interacting with each other in the presence of thermal motion. The interaction energy is called the *force field*in chemistry. The force field is usually obtained by matching experimental data or from quantum mechanics calculations. A large number of force fields have been developed over the time for simulation of macromolecules. CHARMM[23], OPLS[24] and Amber [25]are three popular force fields which are used for protein simulation. In chapter two, we will use atomistic model with CHARMM27 force field as our "exact" model to obtain a coarse-grained model for protein.

In living systems, proteins are solvated in water or other solvents, which must be taken into account in protein modeling and simulation. The solvents produce two interrelated effects: viscous and thermal statistical. These effects form the basis of Langevin dynamics, and their relationship is established by the fluctuation-dissipation theorem[26].

There are two common way to account for the solvent. In *explicit solvent* models [27-30], solvent molecules are explicitly modeled inside and around the protein, typically placed in a periodic cell. Such models are costly as they give rise to large dynamical systems. In contrast, in *implicit solvent* models [31, 32], the solvent is represented as a continuum or simple structured media [33-35].

1.3.2. Coarse-graining

Coarse-graining is an umbrella for model reduction methods whose objective is to bridge the spatial and temporal scales gap between simulation capabilities and reality demands. This is achieved by replacing clusters of atoms with a single particle; the cluster size defines the level of coarse-graining. The clusters can be as small as an atom with attached hydrogen, like in so-called "united atom model"[36, 37], or as large as protein subunits [38, 39]. For proteins, a common cluster is one residue [7, 40, 41].

The central task in developing a coarse-grained model is to identify the effective force field. Many methods for finding the effective force field can be categorized by identifying three major groups of methods [42]:

1. In the inverse methods, parameters of the effective force field are calibrated using experimental, thermodynamic, and/or average structural properties. For example, in the so-called Martini force field [43] the parameters are chosen so that the force field reproduces basic thermodynamic properties.
2. In the multiscale methods, the effective force field is computed directly from the all-atom force field [44-49]. In Chapter 2, we use this approach to obtain a rigorous coarse-grained model for dynamics of proteins.
3. "Minimalistic" models, like go-models [50-52] and elastic network models[7, 53-57], rely on ad hoc assumptions, which make the coarse-graining procedure more efficient, but at the cost of lowering the accuracy. Those assumptions are not always easy to assess *a priori*.

1.3.3. Normal mode analysis

Many protein functions involve large-amplitude collective motion. In terms of the normal mode analysis, such a motion involves low-frequency or slow modes only. Therefore normal mode analysis has been widely used in modeling protein dynamics [58-63].

The principal approximation leading to the normal mode analysis is to replace a rugged potential energy surface with a quadratic surface with a single minimum. With this approximation, the original nonlinear dynamical system is replaced with a linear system whose solution can be constructed using standard methods of linear algebra. The normal mode analysis has been applied to many proteins and processes [64], including

enzyme activities[19, 53] and conformational changes [65, 66]. Of course, there are numerous applications in which nonlinearity plays a crucial role, to which the normal method analysis is inapplicable [67]. In particular, the normal mode analysis is inapplicable to analyzing protein folding.

A typical application of the normal mode analysis to proteins involves the following steps:

1. The initial configuration is obtained from the protein data bank.
2. The initial configuration is used for identifying a local equilibrium state by minimizing the potential energy using one of the standard optimization methods [3].
3. The (Hessian) stiffness matrix is obtained by repeated differentiation of the potential energy at the local minimum.
4. The eigenvalues (frequencies) and eigenvectors (normal modes) are computed for the stiffness matrix normalized by the mass matrix.

These steps exclude viscosity effects, which can be included in the normal mode analysis, provided that one can construct the viscosity (friction) matrix [41, 68, 69]. Furthermore, often protein dynamics can be considered as overdamped[70, 71], so that the inertial effects can be neglected and the eigenvalue analysis is performed for the first rather than second order systems. Accordingly, the eigenvalues are interpreted as relaxation rates rather than frequencies, and the eigenvectors represent the relaxation modes. This dissertation is restricted to such scenarios.

1.3.4. Elastic network models

The first elastic network model was proposed by Tirion in 1996 [72]. The idea was that the low-frequency collective motion is insensitive to atomistic details and can be captured with a highly simplified model. In Tirion's model, the protein is viewed as a truss, whose nodes represent atoms and the bars have constant stiffness; the bars are placed only between nodes separated by a distance less than a cut-off distance. Thus there are only two model parameters – the bar stiffness and the cut-off distance. Later Atilgan et al[7] proposed a coarse-grained elastic network model in which the nodes represented the α-carbons. Another commonly used elastic network model, called Gaussian Network model, was proposed by Bahar and her coworkers in 1998[73]. In this model every node in the network has one degree of freedom, which is the magnitude of displacement.

Over the time, many similar models have been developed to improve these models. Examples are the models which try to improve the accuracy[55, 74-77] and also further coarse-grained models[78-80] to be applied for larger proteins. These models have been applied widely for describing many protein functions involving collective motion. This includes conformational changes of protein upon binding to ligand[81-83], dynamics of allosteric proteins[84], and refinement of X-ray crystallographic experiments[85, 86].

Elastic network models are commonly calibrated by using mean-square atomic displacements (known as B-factors or Debye–Waller factors)[7], obtained by X-ray crystallographic data. However, as we will show in chapter 3, those calibration

procedures ignore the important effect of rigid body motion, which turns out to have a dominant contribution to the B-factors.

1.4. Ideal chain model

One of the simple models for describing elasticity of polymers or proteins in the unfolded configuration is the *ideal chain* or *freely jointed chain*model. This model, despite its simplicity, provides an efficient tool for studying both equilibrium and dynamical properties.

In the model, the chain is viewed as an assembly of rigid links connected to each other by freely rotating joints. It is assumed that the links do not interact with each other. As a result, when the chain is subjected to thermal motion, each link rotates independently of the others. The configuration of an ideal chain is similar to the path traced by random walk[87], and therefore the ideal chain is sometimes called the random walk chain.

To obtain an understanding about the model and its properties, let us consider an ideal chain with N identical links of length b (Fig. 1.3). For the i-th link, we denote the position vectors of the beginning and the end by $\vec{r}_i$and $\vec{r}_{i+1}$, respectively, and define the

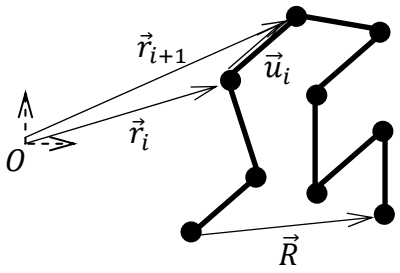


Figure1.3. Schematic of an ideal chain

"bond vector" as $\vec{u}_i = \vec{r}_{i+1} - \vec{r}_i$. Then the end-to-end distance vector $\vec{R}$ can be expressed as

$$\vec{R} = \sum_{i=1}^{N} \vec{u}_i \quad (1.1)$$

The average of the end to end distance $\vec{R}$ is zero, since its projection along any direction has the same probability of having a positive or negative value. The average mean square of $\vec{R}$ is calculated as

$$\langle R^2 \rangle = \langle \sum_{i=1}^{N} \vec{u}_i \cdot \sum_{j=1}^{N} \vec{u}_j \rangle = \sum_{i=1}^{N} \langle \vec{u}_i \cdot \vec{u}_i \rangle + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \langle \vec{u}_i \cdot \vec{u}_j \rangle \quad (1.2)$$

Since the bond vectors are uncorrelated, $\langle \vec{u}_i \cdot \vec{u}_j \rangle = 0$for $i \neq j$, and $\langle \vec{u}_i \cdot \vec{u}_i \rangle = b^2$, one obtains

$$\langle R^2 \rangle = Nb^2 \quad (1.3)$$

An alternative measure of the chain shape is the radius of gyration R_g, defined as[87]

$$\langle R_g^2 \rangle = \frac{1}{2N^2} \langle \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij}^2 \rangle = \frac{Nb^2}{6} \qquad (1.4)$$

where r_{ij} is the distance between ith and jth bead.

Another important quantity for describing the chain is the probability distribution function $P_N(\vec{R})$ that the-end-to-end distance is a prescribed vector $\vec{R}$. For $N \gg 1$, the central limit theorem leads to a Gaussian distribution [87]:

$$P_N(\vec{R}) = \left(\frac{3}{2\pi N b^2}\right)^{3/2} \exp\left[\frac{-3R^2}{2Nb^2}\right] \qquad (1.5)$$

In the model, the source of the potential (strain) energy of the chain is its entropy. Following Boltzmann's construction the entropy S of a macroscopic state is proportional to the logarithm of the number of possible microscopic states corresponding to the macroscopic state [87],

$$S = k_B \ln [\Omega] \qquad (1.6)$$

where k_B is Boltzmann's constant. At a constant temperature T, the free energy A is related to the entropy by the following equation

$$A = -TS = -k_B T \ln[\Omega] \qquad (1.7)$$

For an ideal chain, $\Omega(\vec{R}) \propto P_N(\vec{R})$, and therefore the free energy is calculated as

$$A(\vec{R}) = -k_B T \ln\left[P_N(\vec{R})\right] + C = \mathrm{k_B T} \frac{3R^2}{2Nb^2} + C_1 \qquad (1.8)$$

where$C$ and C_1 are constants independent of $\vec{R}$. From Eq. 1.8, we can derive an expression for the external force $\vec{f}$ which is required to maintain the end-to-end vector $\vec{R}$.

$$\vec{f}(\vec{R}) = \frac{\partial \mathrm{A}(\vec{R})}{\partial \vec{R}} = \frac{3\mathrm{k}_B\mathrm{T}}{Nb^2}\vec{R} \qquad (1.9)$$

Eq. 1.9 is Hooke's law for entropic elasticity, with the stiffness proportional to the temperature and inversely proportional to the square of the mean-square chain length or the square of the radius of gyration.

The classical results presented here are valid for an unconstrained chain. In Chapter 4, we consider the entropic elasticity of a confined polymer chain. In particular, we use similar statistical calculation to derive expressions for the free energy and force for chains confined between two parallel flat plates.

1.5. Objectives and dissertation structure

The overarching goal of this dissertation is to develop and apply coarse-grained models for describing large time-scale dynamics of protein molecules. Such models must combine efficiency, so that they can be useful on large time scales relevant to biochemical processes, and accuracy, so that they account for fine-grain features essential to large time-scale dynamics.

The remainder of the dissertation is organized as follows. In Chapter 2, we present a new rigorous coarse-graining method for describing long-timescale dynamics of

linear systems. The method gives rise to a hierarchy of approximations for coarse-graining and results in effective stiffness, friction, mass and higher order properties for the system. The method reveals that the correct form for the effective friction of the coarse-grained system is non-diagonal, i.e. coarse-grained degrees of freedom have hydrodynamics-like interaction. We show that considering this fact, which is often ignored in *ad hoc* coarse-grained models, is crucial in predicting the dynamics of the original system. The model, then, is applied successfully to studying relaxation dynamics of proteins. The material of Chapter 2 has been published in the Journal of Chemical Physics[88], coauthored with author's supervisor and co-supervisor.

In Chapter 3, we evaluate the calibration procedure in the elastic network models based on fitting to the mean square displacements from x-ray crystallography(B-factors). We show that this calibration is flawed because it does not take into account the rigid body motion of the protein in the crystalline environment. In particular, by proposing an elastic network model that properly takes into account the rigid body motion of proteins, we show that B-factors are dominated by the rigid body motion. This key finding is further proved by fitting B-factors of a data set of 330 proteins to a pure geometrical model which treats the proteins as rigid bodies. The material of Chapter 3 has been published in Physical Biology[89], coauthored with author's supervisor and co-supervisor.

In Chapter 4, we study a problem in which a thin polymer layer bridges two parallel plates. By modeling the polymer molecules as ideal chains, we derive an

analytical expression for force-distance relationship for the two plates for both monodisperse and polydisperse chains.

Chapter Two

Rigorous coarse-graining for the dynamics of linear systems[1]

2.1. Introduction

Many consequential phenomena in molecular biophysics and materials science happen over time and length scales that are inaccessible by fully atomistic computer simulations. So, coarse-graining has become a common strategy for reducing the computational cost involved in such problems. In all coarse-graining schemes, one divides system's degrees of freedom into masters $r_m$and slaves r_s, and eliminates r_s, so that the system is described in terms of $r_m$only. As a consequence, the original system potential $V(r_m, r_s)$ is replaced with an effective coarse-grained potential $V_{CG}(r_m)$. In particular, the V_{CG} that exactly describes the equilibrium statistics of the coarse-grained system is the free energy (also referred to as the potential of mean force) defined by the equation

$$e^{-V_{CG}(r_m)/k_BT} = \int dr_s \, e^{-V(r_m,r_s)/k_BT} \qquad (2.1)$$

Unfortunately, it is usually computationally prohibitive to coarse-grain based on this equation and, in general, it results in system-dependent many-body potentials.

[1] The material of this chapter has been published in 88. Soheilifard, R., D.E. Makarov, and G.J. Rodin, *Rigorous coarse-graining for the dynamics of linear systems with applications to relaxation dynamics in proteins.* Journal of Chemical Physics. **135**(5), coauthored with author's supervisor and co-supervisor.

Therefore, there has been a considerable effort in developing approximate coarse-grained potentials (see [42, 90, 91] and references therein).

Practicality aside, the most severe limitation of Eq. 2.1 is that it only describes equilibrium properties of the system. In contrast, accurate yet practically feasible description of the dynamics of the coarse-grained system remains an open issue. In particular, elimination of the slaves generally results in memory effects in the dynamics of the masters [26, 92-94]. While some efforts have been made in the literature to consider such memory effects systematically (see, e.g.[95, 96]), most routine applications are based upon the phenomenological model [97], in which the time evolution of the master degrees of freedom obeys a Langevin equation governed by the equilibrium potential V_{CG}.

In this chapter, a rigorous hierarchy of approximations is derived for describing coarse-grained dynamics of a system, whose potential is a quadratic function of the atomic displacements. Such systems are common in studies of proteins [63, 67, 98, 99], especially in the contexts of Normal Mode Analysis (NMA) [100-103] and elastic network models [7, 72, 73]. In particular, NMA is often employed for the prediction of conformational changes of proteins [65, 66], refinement of X-ray and NMR data [102, 104-106], and inelastic neutron scattering studies of proteins[107, 108]. Similarly, elastic network models, which consider proteins as collections of beads connected by linear springs, are commonly used for describing conformational fluctuations[7, 72, 73] and

transitions[109-111] in proteins. Furthermore, there is significant interest in coarse-graining of such elastic network models [8, 40, 53, 55, 77, 79, 80, 112-122].

As an application of the method, it is shown how one can construct a one-bead-per-residue coarse-grained system that accurately approximates both the equilibrium properties and the long-timescale part of the relaxation spectrum of the original system, starting from a harmonic approximation applied to an all-atom model of a protein (as in NMA). In doing so, the inadequacy of the more common *ad hoc* coarse-graining approaches is exposed. In particular, it is shown that coarse-graining induces hydrodynamic-like coupling among the masters, even in the absence of hydrodynamic interaction effects caused by the solvent. That is, the assumption that, upon coarse-graining, each bead is subjected to a frictional force that only depends on the bead's velocity is incorrect. Furthermore, it is shown that this effect is considerable for accurately computing the long-timescale part of the relaxation spectrum, which is the focus of many coarse-grained models. In addition, the method reveals that coarse-graining may result in non-diagonal and even negative-definite mass matrix for the master degrees of freedom.

Although our analysis is limited to linear systems, the message regarding inadequacy of *ad hoc* methods is likely applicable to the more general case. In this regard, one could employ the proposed scheme for computing the effective friction and mass matrices, and then combine them with a non-quadratic coarse-grained potential, to obtain an improved description of the coarse-grained dynamics.

The rest of this chapter is organized as follows: Section 2.2 describes the general theory for systematic coarse-graining of a linear system. First the basics of the method is described using a simple model system containing two degrees of freedom and then the governing equations are generalized to systems with arbitrary dimensions. In Section 2.3, the method is used to compute the long-timescale part of the relaxation spectrum for a benchmark set of proteins. Section 2.4 concludes with closing remarks.

2.2. Theory

In this section we show that by eliminating the slaves, which are assumed to be fast compared to the masters, a hierarchy of approximations is obtained. These approximations give rise to effective stiffness, friction, and mass matrices.

2.2.1. Model Problem

To illustrate the method, we start with a simple example which is formulated in the context of deterministic dynamics. The problem involves two coupled harmonic oscillators with a potential energy in the form $V(u_1, u_2) = \frac{1}{2}k_{11}u_1^2 + \frac{1}{2}k_{22}u_2^2 + k_{12}u_1u_2$, where u's are the displacements relative to equilibrium. The oscillators are subjected to velocity-dependent friction forces and obey the following equations of motion:

$$\gamma_{11}\dot{u}_1(t) + k_{11}u_1(t) + k_{12}u_2(t) = f_1(t) \qquad (2.2)$$

$$\gamma_{22}\dot{u}_2(t) + k_{22}u_2(t) + k_{21}u_1(t) = 0 \qquad (2.3)$$

with the initial conditions

$$u_1(0) = u_1^{(0)} \quad \text{and} \quad u_2(0) = u_2^{(0)} \qquad (2.4)$$

Here f_1 is a force applied to the first oscillator and γ's are the friction coefficients. The stiffness coefficient $k_{21} = k_{12}$ but the adopted notation is useful for extending analysis of the model problem to problems with higher dimensionality. Note that Eqs. 2.2 to 2.4 can be solved exactly, by finding the eigenvalues of the system. The system has two eigenvalues, which are either both negative for a constrained system, or a negative and a zero for an unconstrained system. We note that Eqs. 2.2 and 2.3 describe an overdamped system subjected to an arbitrary, generally nonrandom, external force and so their solution does not satisfy the Boltzmann distribution implied by Eq. 2.1. This particular example is however convenient for the purposes of exposing the mathematical structure of the approximations involved in our coarse-graining method.

By solving Eqs. 2.3 and 2.4 for u_2 in terms of u_1, we find:

$$u_2(t) = u_2^{(0)} \exp\left(-\frac{t}{\tau_2}\right) - \gamma_{22}^{-1} k_{21} \int_0^t \exp\left[-\frac{(t-s)}{\tau_2}\right] u_1(s)\mathrm{ds}$$

where $\tau_2 = \gamma_{22}/k_{22}$ is the characteristic relaxation time of the second oscillator. By substituting this equation into Eq. 2.2, an integro-differential equation for u_1 is obtained:

$$\gamma_{11}\dot{u}_1(t) + k_{11}u_1(t) - k_{12}\gamma_{22}^{-1}k_{21}\int_0^t \exp\left[-\tfrac{(t-s)}{\tau_2}\right] u_1(s)\mathrm{ds} = f_1(t) - k_{12}u_2^{(0)} \exp\left(-\tfrac{t}{\tau_2}\right)$$

(2.5)

In terms of this equation, the purpose of coarse-graining is to replace the integro-differential equation with an approximate differential equation. While such an approach is unnecessarily complicated for solving the simple model problem, its usefulness has been well appreciated for larger systems[26].

Let us define the characteristic time for the first oscillator, $\tau_1 = \gamma_{11}/k_{11}$, and assume that $\tau_1 \gg \tau_2$. Under this assumption, we want to approximate Eq. 2.5 for $= O(\tau_1) \gg \tau_2$. This is achieved by doing three simplifications. First the last term in Eq. 2.5 becomes negligible because of the exponential decay. Second, the exponential decay of the kernel of the integral in Eq. 2.5 implies that

$$\int_0^t \exp\left[-\frac{(t-s)}{\tau_2}\right] u_1(s)\mathrm{ds} = \int_0^{t-\tau_1} \exp\left[-\frac{(t-s)}{\tau_2}\right] u_1(s)\mathrm{ds}$$

$$+\int_{t-\tau_1}^t \exp\left[-\frac{(t-s)}{\tau_2}\right] u_1(s)\mathrm{ds} \approx \int_{t-\tau_1}^t \exp\left[-\frac{(t-s)}{\tau_2}\right] u_1(s)\mathrm{ds}$$

These two simplifications are related to short memory of the system for $t = O(\tau_1)$. Finally, since u_1 is a smooth function on the timescale τ_2, it can be expanded into a Taylor series

$$u_1(s) = u_1(t) + \dot{u}_1(t)(s-t) + \frac{1}{2}\ddot{u}_1(t)(s-t)^2 + \cdots \qquad (2.6)$$

with the understanding that $t - s = O(\tau_2) \ll \tau_1$. By retaining only the leading term in this expansion, the approximation for the integral in Eq. 2.5 becomes particularly simple:

$$\int_0^t \exp\left[-\frac{(t-s)}{\tau_2}\right] u_1(s)\mathrm{ds} \approx \tau_2 u_1(t)$$

As a result, the zeroth order approximation of the governing integro-differential equation is obtained in the form of the differential equation

$$\gamma_{11}\dot{u}_1(t) + (k_{11} - k_{12}k_{22}^{-1}k_{21})u_1(t) = f_1(t) \qquad (2.7)$$

Retaining higher order terms in Eq. 2.6 results in the first,

$$(\gamma_{11} + k_{12}k_{22}^{-1}\gamma_{22}k_{22}^{-1}k_{21})\dot{u}_1(t) + (k_{11} - k_{12}k_{22}^{-1}k_{21})u_1(t) = f_1(t) \qquad (2.8)$$

and second order approximation

$$(-k_{12}k_{22}^{-1}\gamma_{22}k_{22}^{-1}\gamma_{22}k_{22}^{-1}k_{21})\ddot{u}_1(t) + (\gamma_{11} + k_{12}k_{22}^{-1}\gamma_{22}k_{22}^{-1}k_{21})\dot{u}_1(t) +$$
$$(k_{11} - k_{12}k_{22}^{-1}k_{21})u_1(t) = f_1(t) \qquad (2.9)$$

of the governing integro-differential equation.

Note that the approximation gives rise to the effective stiffness, friction, and (negative!) mass:

$$\bar{k}_{11} = k_{11} - k_{12}k_{22}^{-1}k_{21} \qquad (2.10a)$$

$$\bar{\gamma}_{11} = \gamma_{11} + k_{12}k_{22}^{-1}\gamma_{22}k_{22}^{-1}k_{21} \qquad (2.10b)$$

$$\bar{m}_{11} = -k_{12}k_{22}^{-1}\gamma_{22}k_{22}^{-1}\gamma_{22}k_{22}^{-1}k_{21} \qquad (2.10c)$$

Here, the effective mass does not take into account the inertial effects, because the original system is considered to be overdamped. These are system properties independent

of the approximation order. In particular, all three approximations involve the same effective stiffness, and the higher order approximations involve the same effective friction, etc. Note that Eq. 2.10a describes a one-dimensional spring with an effective potential $V_{CG}(u_1) = \frac{1}{2}\bar{k}_{11}u_1^2$. It is easy to check, by performing the Gaussian integral over u_2, that this is exactly the coarse-grained potential produced by Eq. 2.1, regardless of the temperature[123]. It should be noted, however, that the problem described by Eqs. 2.2 and 2.3 does not assume thermal equilibrium. In fact, it is the linearity of the problem that causes temperature independence of the coarse-grained potential defined by Eq. 2.1, and its equivalence to the $V_{CG}(u_1)$ found here.

It is straightforward to solve the differential equations associated with the zeroth and first order approximations. However, the differential equation associated with the second order approximation is problematic as it requires an additional initial condition and leads to an unphysical exponential growth mode associated with the negative effective mass. Nevertheless these issues are straightforward to resolve by simply discarding the unphysical mode. Here we emphasize that by going to higher than first order approximation, we improve the approximation to just one mode of the system. Other unphysical modes obtained by higher order approximations do not correspond to the exact system and need to be discarded.

The outlined approximation method results in a hierarchy of coarse-grained models, in which the slow degree of freedom is retained and the fast degree of freedom is

eliminated. From this perspective, the former is regarded as the master and the latter as the slave degrees of freedom.

2.2.2 Multidimensional case

In this section, we extend the approach developed for the two degrees of freedom system to the case of Brownian dynamics of a system with *n* degrees of freedom. The specific scenario considered here is a molecule (e.g., a protein) subjected to an external force. Indeed, there has been significant interest in the viscoelastic behavior of proteins[124], as probed, e.g., in a single-molecule force probe spectroscopy setup[125]. The molecule is in solution: The dynamic effect of the solvent is treated implicitly by introducing effective friction forces which is linearly proportional to the atoms' velocities. Here we further assume that the system is overdamped. That is, the inertial effects are unimportant. This assumption is common in the literature[71, 122, 126]; However we note that the general approach described here can be extended to underdamped and even undamped systems.

Under the above mentioned assumptions, the dynamics of the system is described by the Langevin equation of the form

$$\boldsymbol{\Gamma}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{f}(t) + \tilde{\mathbf{f}}(t) \qquad (2.11)$$

Here $\mathbf{u}(t)$ is the column-vector formed by the displacements relative to the atoms' equilibrium positions,$\boldsymbol{\Gamma}$ and $\mathbf{K}$ are, respectively, the friction matrix and the stiffness

(Hessian) matrix, $\mathbf{f}(t)$is the column-vector formed by the deterministic external forces, and $\tilde{\mathbf{f}}(t)$is a random force vector, which is related to $\boldsymbol{\Gamma}$ via an appropriate fluctuation-dissipation relationship[26]. $\tilde{\mathbf{f}}(t)$ ensures that our system will obey the Boltzmann distribution, when there are no deterministic forces. Since the average value of the fluctuating force is zero, the time evolution of the thermally averaged displacements obeys the same Eq. 2.11 without the random force. That is, thermally averaged displacements obey deterministic equations of motion. In the following, we will always assume the displacements $\mathbf{u}(t)$ to be the thermally averaged values and thus from now on we simply set $\tilde{\mathbf{f}}(t) = 0$. The matrix $\boldsymbol{\Gamma}$ is symmetric and positive definite. The matrix $\mathbf{K}$ is symmetric and positive semi-definite – its six zero eigenvalues are associated with rigid body motion.

In order toextend the approximations developed in Section 2.2.1, we assume that $\mathbf{u}$ is composed of relatively slow $\mathbf{u}_1$and relatively fast $\mathbf{u}_2$ degrees of freedom; the former are regarded as masters and the latter as slaves. We further assume that the external forces are applied to the masters only so that $\mathbf{f}_2(t) = \mathbf{0}$. Indeed, it is reasonable to include the degrees of freedom subjected to the external driving force in the master set. Finally, we assume that the friction matrix is block-diagonal so that it has the form

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{22} \end{bmatrix} \qquad (2.12)$$

Under these assumptions, the governing equations of motion mimic Eqs. 2.2 and 2.3:

$$\boldsymbol{\Gamma}_{11}\dot{\mathbf{u}}_1(t) + \mathbf{K}_{11}\mathbf{u}_1(t) + \mathbf{K}_{12}\mathbf{u}_2(t) = \mathbf{f}_1(t) \qquad (2.13a)$$

$$\boldsymbol{\Gamma}_{22}\dot{\mathbf{u}}_2(t) + \mathbf{K}_{22}\mathbf{u}_2(t) + \mathbf{K}_{21}\mathbf{u}_1(t) = \mathbf{0} \quad (2.13b)$$

and the approximations become:

Zeroth Order: $\boldsymbol{\Gamma}_{11}\dot{\mathbf{u}}_1(t) + \bar{\mathbf{K}}_{11}\mathbf{u}_1(t) = \mathbf{f}_1(t)$ (2.14a)

First Order: $\bar{\boldsymbol{\Gamma}}_{11}\dot{\mathbf{u}}_1(t) + \bar{\mathbf{K}}_{11}\mathbf{u}_1(t) = \mathbf{f}_1(t)$ (2.14b)

Second Order: $\bar{\mathbf{M}}_{11}\ddot{\mathbf{u}}_1(t) + \bar{\boldsymbol{\Gamma}}_{11}\dot{\mathbf{u}}_1(t) + \bar{\mathbf{K}}_{11}\mathbf{u}_1(t) = \mathbf{f}_1(t)$ (2.14c)

where the effective properties mimic Eq. 2.10:

$$\bar{\mathbf{K}}_{11} = \mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21} \quad (2.15a)$$

$$\bar{\boldsymbol{\Gamma}}_{11} = \boldsymbol{\Gamma}_{11} + \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{22}\mathbf{K}_{22}^{-1}\mathbf{K}_{21} \quad (2.15b)$$

$$\bar{\mathbf{M}}_{11} = -\mathbf{K}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{22}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{22}\mathbf{K}_{22}^{-1}\mathbf{K}_{21} \quad (2.15c)$$

The approximations represented by Eq. 2.15 have simple physical interpretation. First, we note that in all approximations the effective potential of the coarse-grained system is governed by the effective stiffness $\bar{\mathbf{K}}_{11}$ and that the force vector remains the same. Moreover, similar to the example of Section 2.2.1, $\bar{\mathbf{K}}_{11}$ is the Hessian matrix of the potential of mean force $V_{\mathrm{CG}}(\mathbf{u}_1)$ obtained from Eq. 2.1. This potential effectively results from the assumption that the slaves follow the masters adiabatically such that $V_{CG}(\mathbf{u}_1) = min_{\mathbf{u}_2} V(\mathbf{u}_1, \mathbf{u}_2)$. Such an assumption has previously been considered by others [39, 70, 122, 123, 127-129]. Conversely, the higher order approximations developed here take into account the effect of the slaves' dynamics on the masters. Specifically, in the first

order approximation, the dynamics of the slaves results in renormalization of the friction forces acting on the masters, which are now described by the new friction matrix $\bar{\mathbf{\Gamma}}_{11}$. In the second order approximation there is a negative-definite effective mass matrix, which leads to *p-6* exponentially decaying, 6 zeros and *p* exponentially growing modes, where *p* is the number of master degrees of freedom; see Appendix. Similar to the model problem of Section 2.2.1, the approximation is constructed by discarding the exponentially growing modes.

In the context of molecular dynamics in solution, the structure of the friction matrix chosen for Eq. 2.12 is somewhat unphysical. In fact, when friction is governed by hydrodynamic resistance of the solvent, it is meaningful to assume that the matrix $\mathbf{\Gamma}$ is either diagonal, rather than-block-diagonal, or dense; a diagonal $\mathbf{\Gamma}$implies that the hydrodynamics interactions are negligible. In this case, coarse-graining qualitatively changes the structure of the friction matrix as, in contrast to $\mathbf{\Gamma}_{11}$, the effective friction matrix $\bar{\mathbf{\Gamma}}_{11}$ is dense. To the best of our knowledge, these effective hydrodynamic interactions, induced by coarse-graining, have not been considered in the relevant literature[41, 70, 97, 122].

When there are hydrodynamic interactions, the matrix $\mathbf{\Gamma}$ is dense [130]. In this case, the off-diagonal blocks make the approximations more complicate, as they introduce higher-order derivatives in comparison to Eqs 2.14. In particular, if $\mathbf{\Gamma}_{12} \neq \mathbf{0}$ and $\mathbf{\Gamma}_{21} \neq \mathbf{0}$, the resulting zeroth and first order approximations are

$$-\boldsymbol{\Gamma}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{21}\ddot{\mathbf{u}}_1(t) + \left[\boldsymbol{\Gamma}_{11} - \boldsymbol{\Gamma}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{21}\right]\dot{\mathbf{u}}_1(t) + \left[\mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}\right]\mathbf{u}_1(t) = \mathbf{f}_1(t) \qquad (2.16a)$$

and

$$\boldsymbol{\Gamma}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{22}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{21}\ddot{\mathbf{u}}_1(t) + \left[-\boldsymbol{\Gamma}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{21} + \boldsymbol{\Gamma}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{22}\mathbf{K}_{22}^{-1}\mathbf{K}_{21} + \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{22}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{21}\right]\ddot{\mathbf{u}}_1(t) + \left[\boldsymbol{\Gamma}_{11} - \boldsymbol{\Gamma}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{21} + \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\boldsymbol{\Gamma}_{22}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}\right]\dot{\mathbf{u}}_1(t) + \left[\mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}\right]\mathbf{u}_1(t) = \mathbf{f}_1(t) \qquad (2.16b)$$

2.2.3. Collective coordinates

In many cases, the system dynamics is better represented in terms of collective coordinates (also referred to as generalized displacements). For example, those may be associated with the center of mass of a subset of atoms rather than with coordinates of individual atoms. In the following, we extend the proposed coarse-graining scheme to such situations.

Suppose that the original displacements **u** are related to the collective coordinates **v** via an invertible linear time-independent map:

$$\mathbf{v}(t) = \mathbf{L}\mathbf{u}(t) \qquad (2.17a)$$

The corresponding generalized forces are then defined by the map

$$\mathbf{g}(t) = (\mathbf{L}^{-1})^{*}\mathbf{f}(t) \qquad (2.17b)$$

where the asterisk denotes matrix transposition. This map follows from the force-displacement conjugacy condition

$$\mathbf{u}(t)^{*}\mathbf{f}(t) = \mathbf{v}(t)^{*}\mathbf{g}(t)$$

Eq. 2.11 can then be rewritten as

$$(\mathbf{L}^{-1})^{*}\boldsymbol{\Gamma}\mathbf{L}^{-1}\dot{\mathbf{v}}(t) + (\mathbf{L}^{-1})^{*}\mathbf{K}\mathbf{L}^{-1}\mathbf{v}(t) = \mathbf{g}(t) \qquad (2.18a)$$

or simply

$$\hat{\boldsymbol{\Gamma}}\dot{\mathbf{v}}(t) + \hat{\mathbf{K}}\mathbf{v}(t) = \mathbf{g}(t) \qquad (2.18b)$$

This equation has the same structure as Eq. 2.11 and it preserves the symmetries of the friction and stiffness matrices. However, upon partitioning the generalized displacements into the masters and slaves, one must either guarantee the condition $\mathbf{g}_2(t) = 0$ or rely on approximations that do not require this restriction.

We emphasize that the outlined transformation from the original to the collective coordinates does not involve coarse-graining since **L** is an invertible map; coarse-graining can be carried out afterwards, if desired. If coarse-graining is carried out by choosing **L** such that the dimension of **v** is less than that of **u**, then the resulting equations of motion would not properly approximate the system dynamics.

2.2.4. Illustrative example

Here, important features of the above approximation method are illustrated by considering the toy model of a linear triatomic molecule as shown in Figure 2.1a. The friction coefficient γ is assumed to be the same for each atom. Furthermore, it is assumed that there are no hydrodynamic interactions. The system is governed by Eq. 2.11 with

$$\boldsymbol{\Gamma} = \gamma \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{K} = k \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1+\alpha & -\alpha \\ 0 & -\alpha & \alpha \end{bmatrix} \text{ with } 0 < \alpha \leq 1$$

We limit our attention to approximate evaluation of the spectral properties and therefore set $\mathbf{f} = \mathbf{0}$. In other words, we are looking for the relaxation modes, which are the solutions of Eq. 2.11 of the form $\mathbf{u}(t)=\mathbf{v}\exp(\lambda t)$. Here λ and $\mathbf{v}$ are, respectively, the eigenvalues and the eigenvectors of the characteristic equation. Note that $-\lambda$ is the relaxation rate and $-1/\lambda$ is the corresponding relaxation time for the eigenmode.

The system has one zero and two negative eigenvalues. The former represents rigid body motion and the latter represent deformation. Let us consider u_1 and u_3 as the master and u_2 as the slave degrees of freedom. Then, using Eq. 2.14, we obtain the following effective properties:

$$\bar{\mathbf{K}}_{11} = k\frac{\alpha}{1+\alpha}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \qquad (2.19a)$$

$$\bar{\boldsymbol{\Gamma}}_{11} = \gamma\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{(1+\alpha)^2}\begin{bmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{bmatrix}\right\} \qquad (2.19b)$$

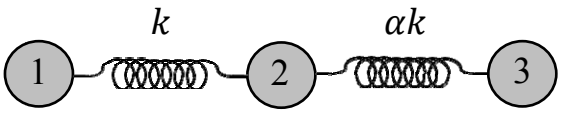
$$\bar{\mathbf{M}}_{11} = -\frac{\gamma^2}{k}\frac{1}{(1+\alpha)^3}\begin{bmatrix}1 & \alpha\\ \alpha & \alpha^2\end{bmatrix} \qquad (2.19c)$$

Note that the zeroth order approximation describes the motion of two particles connected by a spring with a stiffness $k\frac{\alpha}{1+\alpha}$, which, as it should, represents the effective stiffness of the two original springs connected in series. However, in the zeroth order approximation, the correct effective stiffness matrix is combined with an incorrect effective friction matrix. This problem is resolved in the first order approximation. In addition to a renormalization of the diagonal elements in the friction matrix, the correction also induces hydrodynamic-like coupling between the master degrees of freedom, which is represented by the off-diagonal terms. The second order approximation introduces a negative definite effective mass matrix.

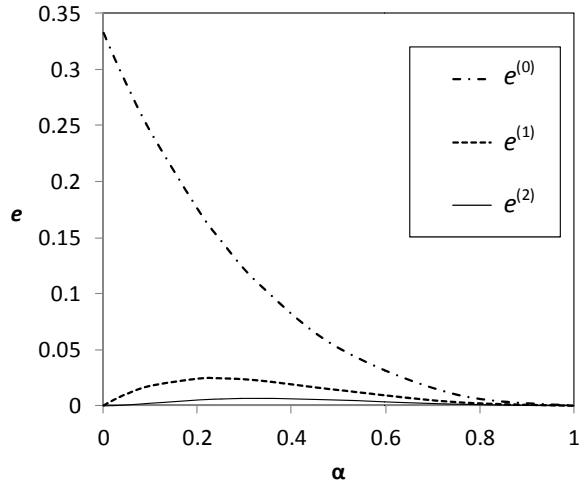
All three approximations involve one zero eigenvalue associated with the rigid body motion mode and one negative eigenvalue associated with the deformation mode. In the following, the latter eigenvalue is used for approximating the slowest deformation mode of the original model. Note that the second order approximation also involves two positive eigenvalues, which result in exponential growth and need to be discarded. To measure the quality of approximations for the eigenvalue λ of the slowest mode, we calculate the error

$$e^{(i)} = \left|\frac{\lambda^{(i)}-\lambda}{\lambda}\right| \qquad (2.20)$$

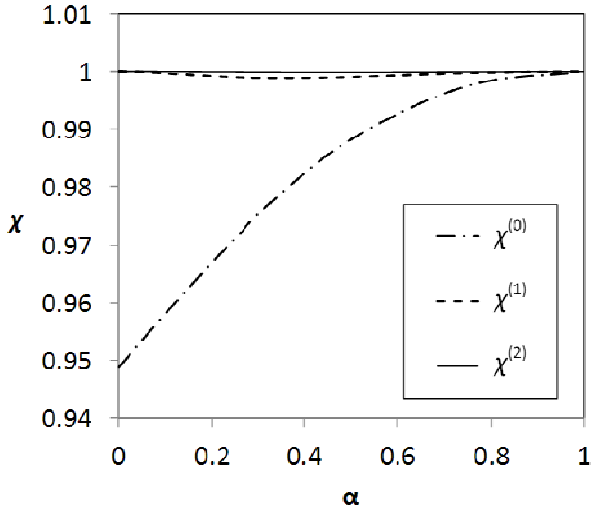
where *i* refers to the approximation order. In Figure 2.1b, the errors are plotted as functions of α. For the symmetric molecule ($\alpha = 1$), all three errors are equal to zero.



2.1a



2.1b



2.1c

Figure 2.1. A case study of coarse-graining: relaxation modes in a toy model of a linear triatomic molecule. (a) The toy model consists of three beads connected by springs and immersed in viscous solvent. The displacement of the middle bead is treated as the "slave" degree of freedom that is to be eliminated by coarse-graining. (b) Errors in the estimate of the slowest relaxation rate for various approximation orders. (c) Correlation coefficient between the exact relaxation eigenmode and that predicted by the coarse-grained model, at different levels of approximation.

However as α approaches zero, only $e^{(1)}$ and $e^{(2)}$approach tend to zero, while $e^{(0)}$approaches 1/3. This qualitative difference between the zeroth and higher order approximations shows the importance of the effective friction matrix. Another interesting and attractive feature of the approximations is that the difference between the zeroth and first order approximations is substantially larger than that between the first and second order approximations. This observation suggests that in most of the cases the first order approximation may be sufficient.

To assess the quality of the approximations $\mathbf{v}^{(i)} = [{v_1}^{(i)}, {v_2}^{(i)}]$ for the eigenvector $\mathbf{v} = [v_1, v_2, v_3]$ of the slowest mode, we should reduce the dimensionality of $\mathbf{v}$ so that both vectors become two-dimensional. The reduced eigenvector $\bar{\mathbf{v}}$ is obtained by (i) suppressing the component v_2, corresponding to the slave degree of freedom, and (ii) normalizing the vector $[v_1, v_3]$ so that the reduced vector has unit length:

$$\bar{\mathbf{v}} = \frac{1}{\sqrt{v_1^2 + v_3^2}} [v_1, v_3]$$

The quality of the approximations is assessed by the correlation coefficient

$$\chi^{(i)} = \left| \bar{\mathbf{v}} \cdot \mathbf{v}^{(i)} \right| \qquad (2.21)$$

Of course $\chi = 1$ represents perfect approximations. It should be noted that for the second and higher order approximations, one must also reduce $\mathbf{v}^{(i)}$by suppressing components associated with exponential growth, followed by normalization.

Figure 2.1c shows the plots for $\chi^{(i)}$as functions of α. For $\alpha = 1$, all three approximations are perfect. As α decreases, $\chi^{(1)}$ and $\chi^{(2)}$ remain close to unity, and approach unity as $\alpha \to 0$; in contrast, $\chi^{(0)} = \frac{3}{\sqrt{10}} \approx 0.949$ as $\alpha \to 0$. Overall, for the chosen example, all three approximations should be regarded as acceptable.

2.3. Applications to relaxation dynamics of proteins

In this Section, the approximations are evaluated through application to a set of protein structures. As before, we limit our attention to predicting the long-timescale part of the relaxation spectrum of the protein and set $\mathbf{f}(t) = \mathbf{0}$ in Eq. 2.11. Furthermore, we seek the relaxation modes in the form $\mathbf{u}(t) = \mathbf{c}\exp(-\lambda t)$, where λ is the relaxation rate and **c** is the corresponding eigenvector.

For each protein, we first construct an all-atom linear model in which **K** is computed as the Hessian matrix of the potential energy in the minimum energy configuration of the molecule [100-103]. The latter is computed by energy minimization starting with the protein databank configuration. In all of the calculations, the TINKER package [131] was used with the Charmm27 force field [23]. Here we report the results obtained using the vacuum force field. This is done to avoid approximation errors in Hessian matrix calculations arising when implicit solvation models are used. Nevertheless, we have also performed the calculations considering several standard

implicit solvation models, and found that the method performance was essentially the same for all those models.

We considered two models for the friction matrix. In the first model, the hydrodynamic interactions were neglected so that $\boldsymbol{\Gamma}$ had only diagonal entries dictated by Stokes' formula:

$$\gamma_i = 6\pi\eta a_i$$

where a_i is the effective hydrodynamic radius and η is the viscosity of water. The effective hydrodynamic radius of each atom was considered to be proportional to its van der Waals radius. We emphasize that our objective here is not to obtain a realistic description of the proteins in question, but rather to evaluate the coarse-graining scheme. The above crude model for the friction coefficients appears to be adequate for this purpose. Indeed, none of our conclusions were found to change when we exploited a presumably more realistic estimate for the effective hydrodynamic radius that is based on the accessible surface area[132].

In the second model, we considered long-range hydrodynamic interactions for the friction matrix. In particular, the friction matrix was computed as the inverse of the mobility matrix whose 3 by 3 blocks are generated by the Rotne-Prager tensor $\boldsymbol{\zeta}$ generalized for non-identical beads[133, 134]. For the *i*-th diagonal block, this tensor is based on Stokes' formula

$$\boldsymbol{\zeta} = (6\pi\eta a_i)^{-1}\mathbf{I} \qquad (2.22a)$$

where **I** is the second rank identity tensor. Here the effective hydrodynamic radius a_i was chosen to be one fourth of its Van der Walls radius. This arbitrary choice was done to avoid atom overlapping. For the off-diagonal blocks, representing the interactions between the *i*-th and *j*-th atoms, the tensor is

$$\boldsymbol{\zeta} = \frac{1}{8\pi\eta d}\left\{(\mathbf{I} + \mathbf{e} \otimes \mathbf{e}) + \frac{a_i^2 + a_j^2}{d^2}\left(\frac{1}{3}\mathbf{I} - \mathbf{e} \otimes \mathbf{e}\right)\right\} \quad (2.22b)$$

where d is the distance between the atoms and **e** is a unit vector directing from the atom *i* to the atom *j*.

Our first example is the trp-cage (pdb code 1L2Y) shown in Figure 2.2. Hydrodynamic interactions were neglected. Following the widely adopted practice [7, 8, 135], we treated the α–carbons (shown in blue) as the masters and the remaining atoms as the slaves. Consequently, the coarse-grained model consisted of 20 atoms. Of the 60 modes, six represent rigid body motion and the remaining 54 represent relaxation dynamics.

We compare the relaxation spectra of the all-atom- and coarse-grained models by computing the approximation errors defined in Eq. 2.20. For the longest relaxation time, the errors were 16, 0.043 and 0.006 for the zeroth, first and second order approximations, respectively. The huge error associated with the zeroth order approximation is not surprising, for in this approximation the friction of the α–carbon atoms does not consider that of the slaves. A reasonable, albeit not rigorously justifiable,

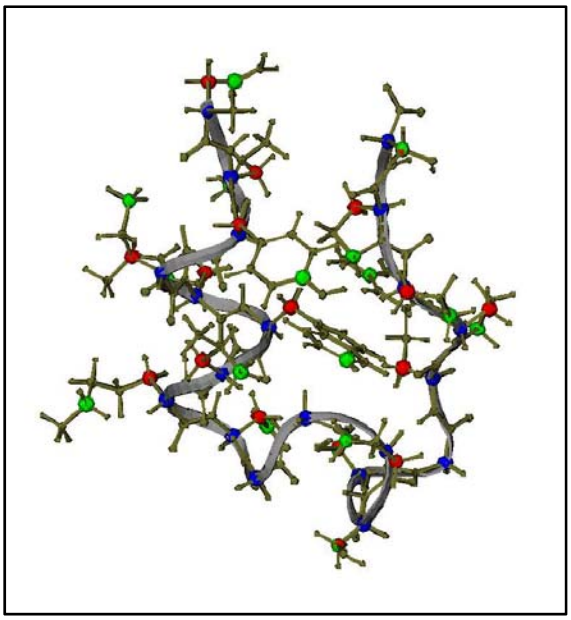


Figure 2.2.Structure of the trp-cage (pdb code 1L2Y). One-bead-per-residue coarse-grained models employed the positions of the α-carbons (blue) as the master degrees of freedom. In addition to the α-carbon, two-bead-per-residue models employed a second carbon to represent the configuration of each amino acid residue except for glycine: This second carbon was chosen to be either the β-carbon (red) or the side-chain carbon separated from the corresponding α-carbon by the maximum number of bonds (green).

approximation is to assume that the total friction coefficient of each α–carbon equals the sum of the friction coefficients of all the atoms forming the residue. This assumption reduces the error from 16 to 0.78. However, this *ad hoc* approximation is still largely inferior to the first and second order approximations, which is based on the correct effective friction matrix. This highlights the importance of the off-diagonal, hydrodynamic-like terms in the coarse-grained friction matrix, which are neglected in the *ad hoc* coarse-grained model. In Figure 2.3 the errors $e^{(0)}$, $e^{(1)}$and $e^{(2)}$are plotted for the 54 slowest relaxation times obtained at various orders of the approximation; the aforementioned *ad hoc* approximation is labeled "$e^{(0m)}$". As seen from the plot, the errors increase as the mode number – and hence the relaxation rate – increases. Going to

higher order further improves the approximation. For example if we limit the errors to 25 percent, then we capture 7 and 11 non-zero relaxation rates using the first and second order, respectively.

Next, we evaluated the performance of the coarse-graining method by comparing the exact and approximate eigenvectors corresponding to the first 20 relaxation modes using the correlation coefficient defined in Eq. 2.21 (Fig. 2.4). The plots reveal that the zeroth order approximation is adequate for the first mode only, whereas the first order approximation accurately predicts the first seven modes. By using the second order approximation the accuracy is improved, but like the first order approximation, it breaks down for the eighth and higher modes.

If it is necessary to capture more relaxation modes, one has to increase the number of masters. Indeed, a one-bead-per-residue model cannot be expected to predict the modes that involve the motion of large side-chains. We thus considered coarse-graining involving two heavy atoms per residue (except for glycine which was still represented by one atom). While the β-carbon (Fig. 2.2, red) seems to be a natural choice for the second atom, this choice did not result in any significant improvement in performance, as compared to the model based on α-carbons only. A better representation of side-chain motion results from considering the second atom as the side-chain carbon that is separated from the α-carbon by the maximum number of bonds, as shown in Fig. 2.2 in green. For the higher-order approximations, this simple extension increased the number

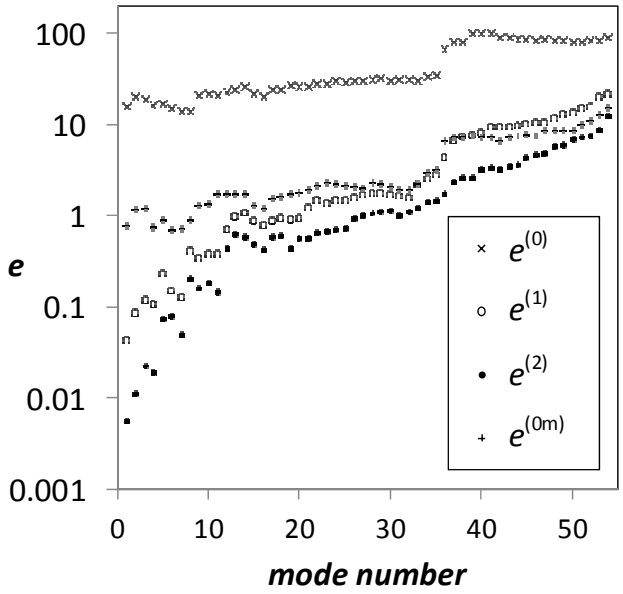


Figure 2.3.One-bead-per-residue model of trp-cage: Errors in the slowest relaxation rates computed for various approximation orders. Here "$e^{(0m)}$" refers to the modified zeroth order model with rescaled friction coefficients. See text for further details.

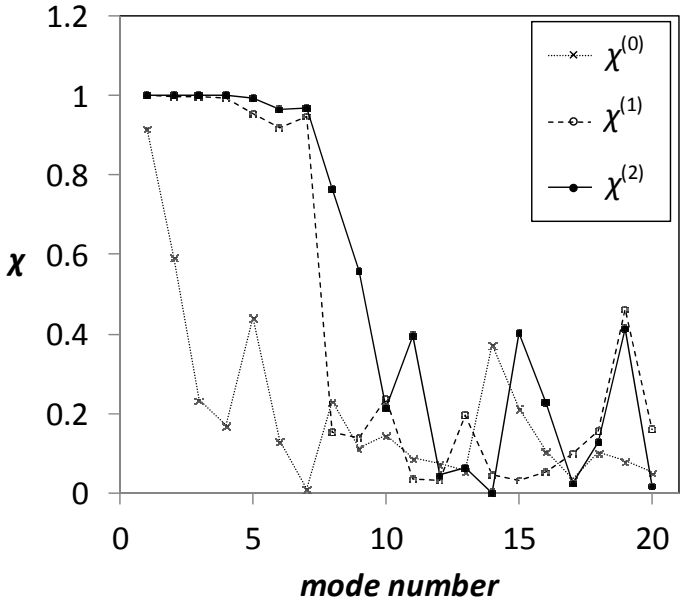


Figure 2.4.One-bead-per-residue model of trp-cage: Correlation coefficients between the exact and approximate relaxation eigenmodes at various approximation orders.

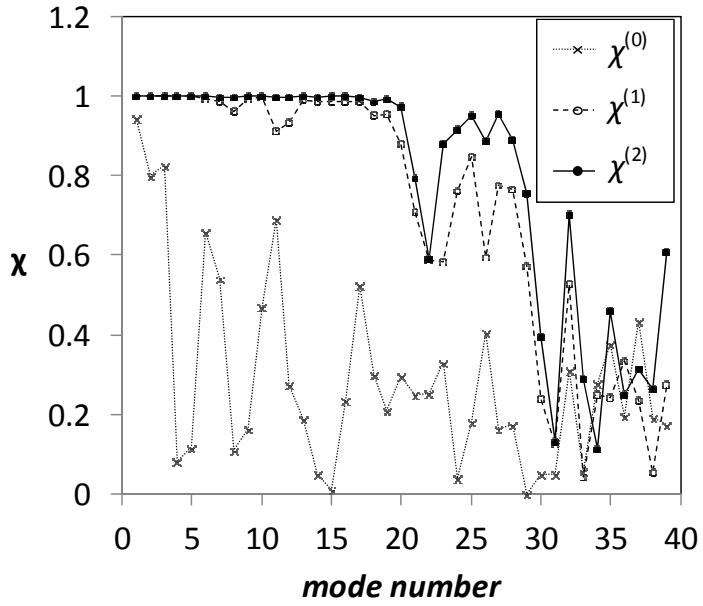


Figure 2.5.Two-beads-per-residue model of trp-cage: Correlation coefficients between the exact and approximate relaxation eigenmodes at various approximation orders.

of captured modes by almost a factor of three (Figure 2.5). We also note that two-bead-per-residue models did little to improve the quality of the zeroth order approximation.

It is also of interest to consider even coarser protein models, in which several residues are represented by a single bead [112, 128, 136]. To this end, one-bead-per-two-residues and one-bead-per-four-residues models are considered, by choosing, respectively, every second and fourth α-carbon as a master degree of freedom. For trp-cage, this resulted in predicting 7 and 4 modes, respectively. This means that going from a one-bead-per-residue to a one-bead-per-two-residues model did not cause loss of accuracy, underscoring potential usefulness of models that represent a group of residues by a single degree of freedom.

We have repeated the calculations for the friction matrix that takes hydrodynamic interactions into account and found only minor changes in the performance of the method. Nevertheless, it is worth noting that in this case our one-bead-per-residue model was able to capture only the first five rather than seven modes.

The observations made for the trp-cage remain valid for other proteins. Table 2.1 summarizes our results for 10 different proteins using one- and two-bead-per-residue models and shows the number of modes that can be captured with a correlation coefficient greater than 0.75 at different orders of approximations. In both models, the number of modes that can be captured with first and second order is an order of magnitude higher than that obtained using the zeroth order approximation. Note that in some cases, using the modified zeroth order approximation in the one-bead-per-residue model increases the number of captured modes as compared to the unmodified zeroth order. Nevertheless, this number is still much lower than that obtained in the first and the second order. Similarly to the trp-cage case, the two-bead-per-residue model significantly outperforms the one-bead-per-residue model, presumably because it captures modes associated with the movement of large side chains.

protein pdb code	number of residues	number of beads per residue	number of captured modes			
			zeroth	modified zeroth	first	second
1L2Y	20	one	1	2	7	8
		two	3	-	24	28
2ERL	40	one	2	2	6	8
		two	7	-	25	29
1P9G	40	one	0	2	4	5
		two	7	-	33	48
1CRN	46	one	2	6	20	22
		two	10	-	22	25
1RB9	52	one	2	3	23	38
		two	11	-	46	46
1IRO	53	one	6	6	20	25
		two	9	-	38	59
2FDN	55	one	1	4	25	26
		two	7	-	37	58
5PTI	58	one	1	2	11	15
		two	10	-	33	34
1C9O	66	one	2	2	17	22
		two	2	-	41	66
1UBQ	76	one	1	6	18	21
		two	14	-	25	31

Table 2.1. Number of relaxation modes accurately captured by various orders of approximation of the one- and two-bead-per-residue coarse-grained models. "Accurately" in this case is defined as having a correlation coefficient greater than 0.75.

In both one- and two-bead-per-residue models, typically, the first order approximation leads to major gains in comparison to the zeroth order approximation, while the additional gains as a result of using the second order approximation are rather modest. Thus, typically, the first order approximation should be sufficient for capturing the first few relaxation modes.

2.4. Concluding remarks

There has been considerable interest in developing reduced-dimensionality, coarse-grained models that accurately represent the dynamics of large molecular systems. While equilibrium statistical-mechanical properties of such models and their connection to their fully atomistic counterparts are well understood, design of *dynamical* models that accurately describe the time evolution of coarse-grained systems remains an open issue. Here we rigorously developed a hierarchy of approximations describing coarse-grained dynamics of linear systems. These approximations are based upon separation between the timescales of the master degrees of freedom (i.e. the ones explicitly included in the coarse-grained model) and the slave degrees of freedom (i.e. the ones eliminated through coarse-graining). In the limit when the characteristic timescales of the slaves are significantly faster than those of the masters, the approximations become exact.

In our method, the coarse-grained dynamics of the master degrees of freedom is described by Langevin-type equations of motion in an effective potential of mean force V_{CG}. This potential is exact, as the equilibrium distribution of the masters is exactly exp(-V_{CG}/k_BT). The coarse-grained equations of motion derived here appear to be similar to those commonly postulated in numerous *ad hoc* coarse-grained models. In contrast to such *ad hoc* models, however, rigorous coarse-graining leads to hydrodynamic-like coupling among the master degrees of freedom. That is, the frictional force acting on a specific master degree of freedom depends not only on its own velocity but also on the velocities of other masters. Our analysis of slow relaxation modes of proteins further

reveals that this dynamic coupling among the master degrees of freedom is essential for capturing relaxation modes and the corresponding relaxation rates. In particular, we have considered the commonly used united-atom model of a protein that uses the positions of its α-carbons as the master degrees of freedom. Under the common assumption that the friction force on each residue (represented by the α-carbon) is proportional to its velocity and is independent of the velocities of other residues, this model fails to correctly capture even a few of the slowest relaxation modes of the proteins. In contrast, the coarse-grained model derived here provides a significantly more accurate description of long-timescale protein dynamics. This finding suggests that standard coarse-grained elastic network models of proteins may be inadequate for describing protein dynamics. This deficiency can however be easily remedied by using the method presented here.

In this study, we have assumed that the dynamics of the underlying atomistic model is overdamped. This assumption, although common, may be too restrictive. Specially, this assumption likely fails for the atoms belonging to the hydrophobic core of a folded protein, which is not exposed to water. We however note that the derivation presented in Section 2.2 can be straightforwardly extended to the cases of undamped and underdamped dynamics. For example, in the undamped case, the second order approximation similar to that described in Section 2.2 results in the equation:

$$\left[\mathbf{M}_{11} + \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{M}_{22}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}\right]\ddot{\mathbf{u}}_1(t) + \left[\mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}\right]\mathbf{u}_1(t) = \mathbf{f}_1(t)$$

Here, we have assumed a block-diagonal structure of the mass matrix, with $\mathbf{M}_{11}$and $\mathbf{M}_{22}$being, respectively, the mass matrices of the master and slave degrees of

freedom. This approximation, which has also been derived in a different way in ref.[137], is particularly useful in the context of normal mode analysis.

Finally, we note that, although our method relies on the linearity of the system, we expect it to be useful in the more general case. Linear approximation is a reasonable starting point for constructing a coarse grained model. In particular, one can employ the non-diagonal friction matrix obtained using the linear approximation, as described here, to construct Langevin equations of motion in a more realistic, nonlinear potential of mean force V_{CG} estimated using standard coarse-graining approaches[42, 90].

Appendix

In this appendix, we address certain technical points associated with the second order approximation given by Eq. 2.14c. In particular, we establish that for negative-definite mass matrix $\bar{\mathbf{M}}_{11}$ the corresponding characteristic equation

$$\mathbf{Q}(\lambda) = \lambda^2\bar{\mathbf{M}}_{11} + \lambda\bar{\mathbf{\Gamma}}_{11} + \bar{\mathbf{K}}_{11} = \mathbf{0} \quad (2.A1)$$

has 6 zero, 3q-6 negative, and 3q positive eigenvalues, where q is the number of master beads.

We start with an observation that the effective stiffness matrix $\bar{\mathbf{K}}_{11}$has 6 zero eigenvalues simply because coarse-graining does not affect rigid body motion modes. So the space of 3q-dimensional vectors can be divided into the six-dimensional null-space $\mathbb{N}(\bar{\mathbf{K}}_{11})$ and (3q-6)-dimensional column space $\mathbb{R}(\bar{\mathbf{K}}_{11})$. For $\mathbf{v} \in \mathbb{N}(\bar{\mathbf{K}}_{11})$ Eq. 2.A1 yields the quadratic equation

$$\mathbf{v}^* Q(\lambda)\mathbf{v} = \lambda^2\mathbf{v}^*\bar{\mathbf{M}}_{11}\mathbf{v} + \lambda\mathbf{v}^*\bar{\mathbf{\Gamma}}_{11}\mathbf{v} = 0$$

and leads to 6 zero and 6 six positive eigenvalues since $\bar{\mathbf{\Gamma}}_{11}$ is positive-definite and $\bar{\mathbf{M}}_{11}$ is negative-definite. For $\mathbf{v} \in \mathbb{R}(\bar{\mathbf{K}}_{11})$ Eq. 2.A1 yields the quadratic equation

$$\mathbf{v}^* Q(\lambda)\mathbf{v} = \lambda^2\mathbf{v}^*\bar{\mathbf{M}}_{11}\mathbf{v} + \lambda\mathbf{v}^*\bar{\mathbf{\Gamma}}_{11}\mathbf{v} + \mathbf{v}^*\bar{\mathbf{K}}_{11}\mathbf{v} = 0$$

whose roots are

$$\lambda_{\pm} = (2\mathbf{v}^*\bar{\mathbf{M}}_{11}\mathbf{v})^{-1}\left(-\mathbf{v}^*\bar{\mathbf{\Gamma}}_{11}\mathbf{v} \pm \sqrt{(\mathbf{v}^*\bar{\mathbf{\Gamma}}_{11}\mathbf{v})^2 - 4(\mathbf{v}^*\bar{\mathbf{K}}_{11}\mathbf{v})(\mathbf{v}^*\bar{\mathbf{M}}_{11}\mathbf{v})}\right)$$

It is clear that this quadratic equation yields (3q-6) positive and (3q-6) negative roots.

Thus if $\bar{\mathbf{M}}_{11}$ is negative-definite, the characteristic equation has 6 zero, 3q-6 negative, and 3q positive eigenvalues. The second order approximation excludes the positive roots as they represent exponentially growing modes.

If $\bar{\mathbf{M}}_{11}$ is negative-semi-definite, the projected quadratic characteristic equation degenerates to a linear one, and as a result the total count of roots is reduced. In practice, this and other computational issues are addressed by replacing the quadratic characteristic equation with a system of linear equations for 6q unknowns. Consequently Eq. 2.A1 is replaced with the linear system

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{M}}_{11} \end{bmatrix} \lambda + \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \bar{\mathbf{K}}_{11} & \bar{\boldsymbol{\Gamma}}_{11} \end{bmatrix} = \mathbf{0}$$

Upon diagonalization of the matrix in front of λ, it becomes clear that a negative-definite $\bar{\mathbf{M}}_{11}$ leads to 6q roots, whereas the number of roots is reduced if $\bar{\mathbf{M}}_{11}$is negative-semi-definite.

Chapter Three

Critical evaluation of simple network models of protein dynamics[2]

3.1. Introduction

Understanding proteins' function requires reliable models describing their internal dynamics. Although molecular dynamics simulations can serve this purpose, they are often computationally prohibitive, especially when long time scales need to be considered[138]. This is definitely the case when slow conformational rearrangements, which are central to many processes involving enzymes and molecular machines, have to be characterized. The need for a manageable alternative to molecular dynamics has lead to the development of coarse grained models, which achieve computational speedup by using collective degrees of freedom and by employing simplified model force fields (see, e.g., ref.[139] for a review).

A natural starting point for a molecule undergoing thermal motion around its equilibrium is to approximate its potential energy as being quadratic. As a result, this approximation has been considerably used to study many aspects of biomolecular dynamics (see [140-144] and references therein). Harmonic models provide simplification in the context of both time-dependent properties[142] and equilibrium

[2] The material of this chapter has been published in 89. Soheilifard, R., D.E. Makarov, and G.J. Rodin, *Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors.* Phys Biol, 2008. **5**(2): p. 026008., coauthored with author's supervisor and co-supervisor.

thermodynamics[145, 146] because in these models the underlying mathematical problem is reduced to solving a system of linear algebraic equations.

The parameters of the harmonic models can be derived from more realistic molecular dynamics force fields. Tirion[147] suggested the fascinating idea that low-frequency vibrational modes of proteins are not sensitive to the fine details of the force field and so they can be predicted by a much cruder model. Tirion's model[147] considers pairwise harmonic interaction among the atoms of a protein, which depends on two parameters, the cut-off radius R and the harmonic bond stiffness γ. It is assumed that two atoms separated by a distance greater than R do not interact while all other pairs of atoms are connected by identical harmonic bonds. The protein is therefore modeled as a network of particles connected by Hookean springs, whose connectivity is described by the cutoff radius. The low-frequency vibrational spectrum and the mean-square atomic displacements (or B-factors) computed with such a simple model agreed incredibly well with calculations that employed a more realistic force field[147].

The two-parameter harmonic potential has subsequently been chosen by several groups [147-157], leading to development of elastic network models (ENMs), also referred to as anisotropic network models (ANMs), and the Gaussian network model[158] or GNM. In this chapter we use the acronym ANM to describe the particular form of ENM as proposed in ref. [148]. The acronym ENM will be reserved for a more general class of elastic network models that will be introduced below. ENM is mathematically equivalent to the classic truss model[159], which is extensively used in

civil engineering. While each network node may correspond to an individual atom, it is currently more common to use coarse-grained networks, in which each node is associated , e.g., with α–carbons of the protein [152].

In GNM [158], each network node has one degree of freedom interpreted as the displacement magnitude. Similar models have been developed in the context of protein folding [153, 160-162] and mechanical protein stretching experiments [163-165]. Recently, GNM, ANM and related models have been applied to various macromolecular systems and have been particularly useful in predicting conformational changes of proteins upon ligand binding and biologically relevant large-amplitude motions (see, e.g., [156, 166-171]).

The appeal of network models is that they promise to capture slow conformational modes of proteins without requiring any knowledge of the specific interactions within a protein. It is only required to have the equilibrium protein geometry and two constants, the cutoff distance R and the bond stiffness γ, which one hopes to be similar for different proteins. This may make ENM particularly appropriate for studies of protein dynamics on the protein databank scale or for very large proteins and their assemblies. Recently, there have been attempts to improve quantitative predictions provided by such simple models [155, 172, 173]. Such efforts are particularly useful in the context of quantifying the mechanical response of proteins in AFM pulling experiments, to which they have been recently applied [163, 174, 175]. Hence obtaining reliable network parameters and

validating models across a large representative set of structures becomes an important issue.

The most common approach for calibrating ANM and GNM is to use the mean-square atomic displacements [154] (known as B-factors or Debye-Waller factors), which are measured by using X-ray crystallography and are available from the Protein Data Bank (www.rcsb.org). Molecular dynamics simulations have also been employed to parametrize an ENM [176, 177] but this procedure gives rise to a model that is not transferrable to other systems so that part of the appeal of ENM is lost.

The typical correlation between crystallographic B-factors and ANM or GNM is only modest [154, 155] (see the Results section for further discussion of this issue). However the fact that ANM or GNM is capable of capturing a set of n mean-square displacements (where n corresponds to the number of residues, typically of order of one hundred or more) using only two fitting parameters is definitely not accidental. As a result, B-factor-parametrized ANM and GNM have become widely accepted [178].

However, as it is well known in statistics, a correlation between a model and data does not necessarily mean a causal relationship. Halle, for example, has suggested an alternative model, which correlates the B-factor for each atom with the number of its non-covalently bonded neighbors[179]. In Halle's model, the mean square displacement of an atom is vastly determined by its proximity to the surface of the protein. This highlights the possibility pointed out by Thorpe [149, 157] that GNM (or ANM) may correlate with the crystallographic B-factors simply because it correlates with another,

more essential property determining the B-factors (which is the proximity to the surface according to Halle's model). In this work we analyze the B-factors for a set of 330 crystallographic structures that is composed of data sets previously used to test and validate various simple models of protein dynamics including ANM, GNM, and the Halle model. According to our analysis, we argue that rigid body motion, which is not commonly accounted for in ANM and GNM, is the primary contributor to the crystallographic B-factors. Thus the principal message of this work is twofold:

1. Calibration procedures commonly used by ANM and GNM are flawed because they do not properly account for rigid body motion and constraints imposed on the protein by the surrounding crystal.

2. Rigid body motion is the dominant part of the crystallographic B-factors and so the latter are poorly suited for extracting intrinsic elastic properties of proteins.

It should be noted that we do not challenge the usefulness of ENMs *per se*. Rather we expose essential flaws in the way some of them are calibrated and question the usefulness of the crystallographic B-factors for identifying network properties.

The idea that protein motions in crystals are dominated by rigid body motion is not new. It goes back forty years to the translation libration screw (TLS) model [180], which was considered for modeling protein dynamics by Sternberg and co-workers [181] and later by Kuriyan and Weis [182]. TLS correlates well with the X-ray data for a small set of proteins studied in ref. [182]. Kundu, Melton, Sorensen, and Phillips (KMSP) [183] suggested an *ad hoc* simplification of the TLS model in which the center of rotation

coincides with the protein's center of mass. Based on their fit of the B-factors for 113 proteins, Kundu and co-workers concluded that their model is inferior to GNM. Song and Jernigan [155] proposed their vGNM by enriching the KMSP model with degrees of freedom representing the elastic deformation. The enrichment employs the eigenvectors of GNM. Song and Jernigan [155] further used the set of the 113 proteins previously used by Kundu and co-workers [183] to show that vGNM provides a superior fit to the experimental B-factors as compared to the KMSP model and to GNM. To a certain extent, this is not surprising because, for the set of proteins chosen by Kundu and co-workers, a typical correlation coefficient for GNM is only about 0.6 and vGNM must always outperform the KMSP model simply because it has more data fitting flexibility. Kidera, Go, and co-workers [184-186] have developed a method for refinement of protein structure that correctly takes into account both internal motion in the harmonic approximation and rigid body motion within the TLS model. In principle their method permits one to separate the two contributions to the crystallographic B-factors. So far this approach has been applied only to a few proteins. Practically, their method needs prior knowledge of the underlying atomistic interaction potential used to perform initial normal mode analysis, in contrast to the ENM approach that strives to derive an effective harmonic potential directly from X-ray data.

In this chapter, we develop two modifications of the TLS model. In the first modification, referred to as rTLS, we reduce the number of fitting constants by choosing the center of rotation prior to computing the fitting parameters. Opposite to the KMSP model, the center of rotation is located so that the prediction of the rTLS and TLS would

coincide if the models were applied to an ideal continuous rigid body. The eTLS model enhances TLS by allowing for flexible chain ends. It is shown that eTLS considerably outperforms GNM, ANM, vGNM, and KMSP models when applied to the proteins in the data sets that were previously used to validate and test those models, proposing that rigid-body motion tends to dominate the B-factors for those proteins, making them inappropriate for calibration of network models.

It is important to notice that there are other reasons why the utility of crystallographic B-factors for developing network models may be restricted [185]. Firstly, the B-factors do not distinguish between static disorder and dynamic fluctuations, while network models only deal with the latter. Secondly, B-factors are not truly "experimental" because they are obtained through a refinement procedure that fits X-ray diffraction data to a model. The second limitation is unavoidable if one only uses the information available from the protein databank[184, 185]. The above caveats only further strengthen the conclusion of this work that B-factors are poorly suited for calibrating network models of protein dynamics.

Here, we will consider the same assumptions about the data as chosen both in previous ENM studies [148, 149, 151, 152, 154, 155, 171, 178] and in most competing models[155, 179, 182, 183]; we will therefore suppose that the B-factors from the protein databank represent "true" mean square displacements due to proteins' dynamics.

In the remainder of this chapter we will explain and compare two types of theories for protein motion in crystals. First, we will develop a general theory for elastic network

models (ENMs) which includes the effects of a protein's rigid body motion and the constraints imposed by the confining crystal. This development results in a proper relationship between the network stiffness matrix and B-factors, which is different from that used in ANM and GNM. Secondly, we will explain the TLS model and its modifications. The physical assumptions adopted by ANM, GNM, TLS and their modifications are mutually conflicting, yet different data sets were used in support of different models, so that no consensus exists in the literature as to what model better captures the physics. To settle this matter, we will present a comparison of various competing simple models of protein dynamics in crystals using all available data sets previously used.

3.2. Methods

3.2.1. Elastic network models

A. Intrinsic and extrinsic protein stiffness matrices

In elastic network models for protein dynamics, the effective elastic energy of a deformed protein is assumed to be a quadratic function of the atomic displacements,

$$E = \frac{1}{2}\mathbf{u}^T\mathbf{K}\mathbf{u} \qquad (3.1)$$

where $\mathbf{u}$ is the column-vector containing atomic displacements and $\mathbf{K}$ is the network stiffness matrix, which is completely determined by the network geometry and its elastic properties. As noted above, it is a common practice to use the positions of the C_α atoms

as the network nodes, although all-atom networks have also been considered[147]. For ENM, the dimension of $\mathbf{u}$ is 3*n*, where *n* corresponds to the number of nodes (C_α atoms). Elastic networks have been analyzed in depth and the basic results presented in this section are well established in the literature[159]. For GNM the dimension of $\mathbf{u}$ is *n*.

The matrix $\mathbf{K}$ is symmetric semi-positive definite which has at least six zero eigenvalues. If there are precisely six zero eigenvalues, the null space $\mathsf{N}(\mathbf{K})$ is spanned by vectors representing three rigid translations and three rigid rotations, so that any $\mathbf{u}$ related to deformation, $\mathbf{u} \notin \mathsf{N}(\mathbf{K})$, leads to $E > 0$. Such networks are termed stable. An example of a two-dimensional stable network is one where the bonds form a triangle. On the contrary, if the edges form a square, the network is unstable. Recent work proposes that protein unfolding can be viewed as loss of network stability[187]. In what follows we consider only stable networks. Note that $\dim \mathsf{N}(\mathbf{K}) = 1$ for GNM [157]. Thus the GNM setting is not sufficient for a meaningful treatment of rigid body motion. For this reason, we do not consider the GNM in this section. Nevertheless GNM will be included in the comparisons presented below.

For a stable network, by the spectral decomposition of $\mathbf{K}$ we obtain the following expression:

$$E = \frac{1}{2}\mathbf{u}^T\mathbf{K}\mathbf{u} = \frac{1}{2}\mathbf{u}^T\mathbf{R}^T\mathbf{\Lambda}\mathbf{R}\mathbf{u} = \frac{1}{2}\mathbf{v}^T\mathbf{\Lambda}\mathbf{v} = \frac{1}{2}\sum_{\alpha=1}^{3n}\lambda^\alpha v_\alpha^2 = \frac{1}{2}\sum_{\alpha=7}^{3n}\lambda^\alpha v_\alpha^2 \qquad (3.2)$$

where $\mathbf{R}$ is the orthogonal matrix formed by the eigenvectors of $\mathbf{K}$, $\mathbf{\Lambda}$ is the diagonal matrix containing the eigenvalues λ^α of $\mathbf{K}$, and $\mathbf{v} = \mathbf{Ru}$. The equipartitioning theorem implies

$$\lambda^\alpha \left\langle v_\alpha^2 \right\rangle = k_B T \text{ for } \alpha > 6 \tag{3.3}$$

and

$$\left\langle v_\alpha v_\beta \right\rangle = 0 \text{ for all } \alpha \neq \beta \tag{3.4}$$

where the angular brackets denote time or ensemble averaging. Note that Eqs. 3.3-4 do not distinguish between two network trajectories that differ by rigid body motion only.

Commonly rigid body motion is uniquely defined by geometric constraints imposed on the network (displacement boundary conditions). Then the stiffness matrix is modified by removing the rows and columns associated with the constrained degrees of freedom. Eq. 3.1 must then be augmented by including the constraints. If geometric constraints are imposed on six degrees of freedom only, so that rigid body motion is fixed but the deformation is unaffected, the formulation involves a $(3n-6)\times(3n-6)$ positive-definite stiffness matrix $\hat{\mathbf{K}}$ and the elastic energy is represented by

$$\hat{E} = \frac{1}{2}\hat{\mathbf{u}}^T \hat{\mathbf{K}} \hat{\mathbf{u}} \tag{3.5}$$

where the hat refers to the geometrically constrained quantities. Notice that the eigenvalues of $\hat{\mathbf{K}}$ are intrinsic network properties, whereas the eigenvectors rely on a particular choice of geometric constraints.

If the network represents a protein embedded within a crystalline sample, we can decompose the stiffness matrix of the entire network as

$$\hat{\mathbf{K}} = \begin{bmatrix} \mathbf{K}_{pp} & \hat{\mathbf{K}}_{pc} \\ \hat{\mathbf{K}}_{cp} & \hat{\mathbf{K}}_{cc} \end{bmatrix} \tag{3.6}$$

where the subscripts *p* and *c* refer to the protein and the sample, respectively. The absence of the hat symbol over $\mathbf{K}_{pp}$ implies that the protein atoms are not geometrically constrained. Now we can write the elastic energy of the protein in the form

$$E_p = \frac{1}{2}\mathbf{u}_p^T \mathbf{K}^e_{pp} \mathbf{u}_p \tag{3.7}$$

where

$$\mathbf{K}^e_{pp} = \mathbf{K}_{pp} - \hat{\mathbf{K}}_{pc}\hat{\mathbf{K}}^{-1}_{cc}\hat{\mathbf{K}}_{cp} \tag{3.8}$$

and the superscript *e* is used to emphasize that $\mathbf{K}^e_{pp}$ is an *extrinsic* (or apparent) property of the protein. The matrix $\mathbf{K}^e_{pp}$ is directly associated with the mean-square atomic displacements or B-factors obtained from X-ray crystallography. However, in order to construct $\mathbf{K}^e_{pp}$ we need the stiffness matrix $\hat{\mathbf{K}}$ of the entire crystal.

The *intrinsic* stiffness matrix $\mathbf{K}^i_{pp}$ describes the protein decoupled from its surroundings. It is acquired following the standard rules for constructing network stiffness matrices, see, e.g., ref.[148] for details. Its entries are related only to the stiffness

of the effective harmonic bonds connecting atoms within the protein. As a result it is easy to construct using the protein structural information[148]. Like the matrix **K**, the matrix $\mathbf{K}_{pp}^{i}$ is a property of an unconstrained network and so its null-space contains rigid body motion. Also note that $\mathbf{K}_{pp}^{i} \neq \mathbf{K}_{pp}$ because $\mathbf{K}_{pp}$ is affected by the stiffness of the bonds connecting the atoms of the protein to the atoms of the surrounding crystal. In contrast to the extrinsic stiffness matrix $\mathbf{K}_{pp}^{e}$, the intrinsic stiffness matrix $\mathbf{K}_{pp}^{i}$ is easy to construct but difficult to relate to the experimental B-factors.

The fact that B-factors are not intrinsic protein properties has been highlighted in ref. [188], which shows their important dependence on crystal packing. In a different context, the importance of rigid body motion has been recognized in the work of Kidera and Go[185, 186].

B. Extrinsic stiffness matrix versus B-factors

Consider the covariance matrix of the atomic displacements [145, 146],

$$\mathbf{C} = \left\langle \mathbf{u}_p \otimes \mathbf{u}_p \right\rangle, \tag{3.9}$$

The mean-square displacements (B-factors) are related to the diagonal elements of the matrix by

$$B^{\alpha} = \frac{8\pi^2}{3}(C^{3\alpha-2,3\alpha-2} + C^{3\alpha-1,3\alpha-1} + C^{3\alpha,3\alpha}) \tag{3.10}$$

Here, the index $\alpha = 1, \ldots, n_p$ enumerates the atoms that serve as the network nodes. On the other hand, using the elastic energy from Eq. 3.7, the equipartitioning theorem results in the following relationship between the covariance matrix and the stiffness matrix:

$$\mathbf{C} = k_B T \left(\mathbf{K}^e_{pp} \right)^{-1} \tag{3.11}$$

Thus if the number of nodes representing a protein is denoted by n_p, there are n_p simple algebraic relationships involving the diagonal elements of the extrinsic compliance matrix $\left(\mathbf{K}^e_{pp} \right)^{-1}$ and the B-factors.

It is common to relate the properties of an ANM to the experimentally measured atomic fluctuations by using the relationship [148, 154]:

$$\mathbf{C} = k_B T \left(\mathbf{K}^i_{pp} \right)^{-1} \tag{3.11'}$$

Since the matrix $\mathbf{K}^i_{pp}$ is singular, a pseudoinverse is used to construct $\left(\mathbf{K}^i_{pp} \right)^{-1}$. Use of a pseudoinverse sweeps the physical problem with Eq. 3.11' under the rug. Eq. 3.11' cannot be correct because $\mathbf{C}$ is affected by how the protein is constrained within a crystal, whereas $\mathbf{K}^i_{pp}$ is not. The covariance matrix $\mathbf{C}$ (or B-factors) of an unconstrained protein is not well defined because it is not invariant under rigid-body motion and hence it depends on the reference frame[146]. Moreover, in contrast to Eq. 3.11' relating a positive definite matrix to a semi-positive definite matrix, Eq. 3.11 relates two positive definite matrices.

C. Intrinsic stiffness matrix versus B-factors

Since the construction the extrinsic stiffness matrix $\mathbf{K}^e_{pp}$ requires the stiffness matrix of the entire crystalline sample, the use of the proper relationship between $\mathbf{K}^e_{pp}$ and the B-factors is impractical. Here we construct an approximate relationship between $\mathbf{K}^i_{pp}$ and the B-factors assuming that the difference $\mathbf{\Delta} = \mathbf{K}^e_{pp} - \mathbf{K}^i_{pp}$ is small.

In the relationship

$$\mathbf{C} = k_B T \left(\mathbf{K}^e_{pp} \right)^{-1} = k_B T \left(\mathbf{K}^i_{pp} + \mathbf{\Delta} \right)^{-1} \tag{3.12}$$

even a small $\mathbf{\Delta}$ considerably affects the null-space of $\mathbf{K}^i_{pp}$. The simplest approximation that emphasizes this fact is

$$\Delta = \sum_{\alpha=1}^{6} \rho^\alpha \mathbf{e}^\alpha \otimes \mathbf{e}^\alpha \tag{3.13}$$

where $\mathbf{e}^\alpha$ are the eigenvectors of $\mathbf{K}^i_{pp}$, the first six of which span the null-space of this matrix. This form represents a weakest crystalline constraint, which is adequate for restraining rigid body motion but does not affect the intrinsic protein stiffness. Physically, Eq. 3.13 implies that the only effect of the surrounding crystal on the protein is to impose an additional harmonic confinement potential along the collective degrees of freedom associated with the rigid-body motions.

For the chosen $\mathbf{\Delta}$, $\mathbf{K}_{pp}^{i}$ and $\mathbf{K}_{pp}^{e}$ have the same eigenvectors and $3n_p - 6$ eigenvalues. This results in the relationship

$$\sigma^{\alpha} = k_B T \begin{cases} \dfrac{1}{\rho^{\alpha}} & \alpha \le 6 \\ \dfrac{1}{\mu^{\alpha}} & \alpha > 6 \end{cases} \tag{3.14}$$

in which σ and μ are the eigenvalues of $\mathbf{C}$ and $\mathbf{K}_{pp}^{i}$, respectively. Note that while ρ^{α} represents the constraint stiffness, the terms $k_B T / \rho^{\alpha}$ correspond to the covariance matrix for rigid body motion. Indeed, the relationship

$$\mathbf{C}' = \mathbf{C} - \sum_{\alpha=1}^{6} \frac{k_B T}{\rho^{\alpha}} \mathbf{e}^{\alpha} \otimes \mathbf{e}^{\alpha} = \sum_{\alpha=1}^{3n_p} \sigma^{\alpha} \mathbf{e}^{\alpha} \otimes \mathbf{e}^{\alpha} - \sum_{\alpha=1}^{6} \frac{k_B T}{\rho^{\alpha}} \mathbf{e}^{\alpha} \otimes \mathbf{e}^{\alpha} = k_B T \sum_{\alpha=7}^{3n_p} \frac{1}{\mu^{\alpha}} \mathbf{e}^{\alpha} \otimes \mathbf{e}^{\alpha} \tag{3.15}$$

shows that the intrinsic compliance matrix is related to the reduced correlation matrix $\mathbf{C}'$ rather than $\mathbf{C}$. Of course the current construction of the matrix $\mathbf{C}'$ is valid only if $\mathbf{\Delta}$ has the form represented by Eq. 3.13.

The fact that $\mathbf{K}_{pp}^{i}$ is directly related to $\mathbf{C}'$ rather than $\mathbf{C}$ can be acknowledged without any computations: $\mathbf{K}_{pp}^{i}$ is invariant under rigid body motion. This means that the covariance matrix to which it is directly related must also be invariant under rigid body motion. Since the covariance matrix $\mathbf{C}$ is not invariant under rigid body motion[146], the relationship between $\mathbf{C}$ and $\mathbf{K}_{pp}^{i}$ must involve only the component of $\mathbf{C}$ that is invariant under rigid body motion. In general, the extraction of this component can be complicated.

Kidera and Go[185, 186] developed a refinement procedure for extracting $\mathbf{K}_{pp}^{i}$ from crystallographic data. In practice, their procedure requires an initial guess for $\mathbf{K}_{pp}^{i}$ that is based on a more elaborate protein force field. Zeng et al [189] has proposed an extraction procedure in the context of molecular dynamics simulations.

3.2.2. TLS models

So far our emphasis was on extracting the component of the covariance matrix **C** that is related to the intrinsic stiffness matrix. In this section, we take a diametrically opposite view and focus on rigid body motion as the primary contributor to **C**.

A. The standard TLS model

Let us assume that rigid body motion is infinitesimal. Then the displacement vector can be prescribed in the form

$$\boldsymbol{u}(\boldsymbol{x},t) = \boldsymbol{u}(\boldsymbol{c},t) + \boldsymbol{\omega}(t) \times (\boldsymbol{x} - \boldsymbol{c}) \tag{3.16}$$

Here, $\boldsymbol{c}$ is the position of a chosen reference point and $\boldsymbol{x}$ is the position of an atom in the equilibrium configuration. The boldface Italics denote true vectors in $\mathbb{R}^3$ rather than column-vectors denoted by the boldface Romans. Any point $\boldsymbol{c}$ can serve as the reference point, since

$$\boldsymbol{u}(\boldsymbol{x},t)=\boldsymbol{u}(\boldsymbol{c},t)+\boldsymbol{\omega}(t)\times(\boldsymbol{x}\text{-}\boldsymbol{c})=\boldsymbol{u}(\boldsymbol{c}',t)+\boldsymbol{\omega}(t)\times(\boldsymbol{x}\text{-}\boldsymbol{c}') \qquad (3.16')$$

for any $\boldsymbol{c}'$. Particularly, if $\boldsymbol{u}(\boldsymbol{c},t)=0$ then $\boldsymbol{c}$ is the center of rotation and the motion becomes pure rotation with respect to $\boldsymbol{c}$. Later on, we identify a particular choice of $\boldsymbol{c}$ that is convenient for our purposes

The mean square displacement of an atom is represented by

$$B(\boldsymbol{x})=\langle\tilde{\boldsymbol{u}}\cdot\tilde{\boldsymbol{u}}\rangle=\langle\tilde{\boldsymbol{v}}\cdot\tilde{\boldsymbol{v}}\rangle+2\langle\tilde{\boldsymbol{v}}\cdot[\tilde{\boldsymbol{\omega}}\times(\boldsymbol{x}-\boldsymbol{c})]\rangle+\langle[\tilde{\boldsymbol{\omega}}\times(\boldsymbol{x}-\boldsymbol{c})]\cdot[\tilde{\boldsymbol{\omega}}\times(\boldsymbol{x}-\boldsymbol{c})]\rangle \qquad (3.17)$$

where the tilde refers to the fluctuation relative to the time average (e.g., $\tilde{\boldsymbol{u}}=\boldsymbol{u}-\langle\boldsymbol{u}\rangle$) and $\boldsymbol{v}=\boldsymbol{u}(\boldsymbol{c},t)$. This expression can be rewritten in the form

$$B(\boldsymbol{x})=\langle\tilde{\boldsymbol{v}}\cdot\tilde{\boldsymbol{v}}\rangle+2\langle\tilde{\boldsymbol{v}}\times\tilde{\boldsymbol{\omega}}\rangle\cdot(\boldsymbol{x}-\boldsymbol{c})+(\boldsymbol{x}-\boldsymbol{c})\cdot\langle\boldsymbol{\Omega}\rangle\cdot(\boldsymbol{x}-\boldsymbol{c}) \qquad (3.18)$$

where $\boldsymbol{\Omega}=\tilde{\boldsymbol{\omega}}\cdot\tilde{\boldsymbol{\omega}}\mathbf{I}-\tilde{\boldsymbol{\omega}}\otimes\tilde{\boldsymbol{\omega}}$ and $\mathbf{I}$ denotes the second rank identity tensor. This form emphasizes that the B-factors are described by a quadratic form whose coefficients are the scalar $\langle\tilde{\boldsymbol{v}}\cdot\tilde{\boldsymbol{v}}\rangle$, vector $\langle\tilde{\boldsymbol{v}}\times\tilde{\boldsymbol{\omega}}\rangle$, and the symmetric second rank tensor $\langle\boldsymbol{\Omega}\rangle$. Note that the special structure of the tensor $\boldsymbol{\Omega}(t)$, which allows one to reconstruct the entire tensor from the vector $\tilde{\omega}(t)$, is fragile under time-averaging. For example, the tensor $\boldsymbol{\Omega}(t)$ has one zero and two equal positive eigenvalues. This eigenvalue structure is preserved by $\langle\boldsymbol{\Omega}\rangle$ if and only if $\tilde{\boldsymbol{\omega}}(t)/|\tilde{\boldsymbol{\omega}}(t)|$ is independent of time. Conversely, the symmetry property of $\boldsymbol{\Omega}(t)$ is preserved under time-averaging. Thus we treat $\langle\boldsymbol{\Omega}\rangle$ as a symmetric positive definite tensor. We note that one of the criticisms of the TLS model in the form that was

previously used [182] was that the property of the rotation matrix to be semi-positive definite was not enforced, leading to potentially physically meaningless results [155]. In the present study this property was explicitly enforced.

In the TLS model [180], $\boldsymbol{c} = 0$ and $\langle \tilde{\boldsymbol{v}} \cdot \tilde{\boldsymbol{v}} \rangle$, $\langle \tilde{\boldsymbol{v}} \times \tilde{\boldsymbol{\omega}} \rangle$, and $\langle \boldsymbol{\Omega} \rangle$ are treated as fitting parameters, so that the total number of fitting parameters is equal to ten. The fitting procedure can be such that the constraint that $\langle \boldsymbol{\Omega} \rangle$ is positive definite is enforced *a priori*. That is, one can determine the fitting parameters using constrained, rather than unconstrained, optimization. Here, we relied on the constrained optimization procedure implemented in the *Mathematica* program. Below we propose two modifications of the TLS model.

B. Reduced TLS model (rTLS)

Let us choose

$$\boldsymbol{c} = \langle \boldsymbol{\Omega} \rangle^{-1} \cdot \langle \tilde{\boldsymbol{v}} \times \tilde{\boldsymbol{\omega}} \rangle \tag{3.19}$$

Then Eq. 3.18 becomes:

$$B(\boldsymbol{x}) = B(\boldsymbol{0}) + \boldsymbol{x} \cdot \langle \boldsymbol{\Omega} \rangle \cdot \boldsymbol{x} = B_{\min} + \boldsymbol{x} \cdot \langle \boldsymbol{\Omega} \rangle \cdot \boldsymbol{x} \tag{3.20}$$

This form implies that the reference point $\boldsymbol{c}$ given by Eq. 3.19 is the best choice of the center of rotation. If we ignore the fact that our model is a collection of discrete atoms

(nodes), then this center of rotation should coincide with the location of the atom that has the smallest B-factor, so that one can fit the data $B(\boldsymbol{x})-B_{\min}$ with a quadratic form. Consequently, the new fit involves $n-1$ data points and six constants, as opposed to *n* data points and ten constants in the TLS model. The difference between the two fitting procedures can be explained in geometrical terms. Both procedures attempt to fit the experimental data with a paraballoid in the four-dimensional space spanned by $\boldsymbol{x}$ and *B*. In the TLS model, $\boldsymbol{c}$ is chosen at the origin, and the scalar-, vector-, and tensor-valued fitting parameters are used to construct an optimal paraballoid; this includes its shape, position, and orientation. In the rTLS, the location of $\boldsymbol{c}$ is defined via the equation $B(\boldsymbol{c})=B_{\min}$. This equation also implies $\langle\tilde{\boldsymbol{v}}.\tilde{\boldsymbol{v}}\rangle=B_{\min}$ and eliminates the vector-valued fitting parameter $\langle\tilde{\boldsymbol{v}}\times\tilde{\boldsymbol{\omega}}\rangle$. As a result the total number of fitting parameters in the rTLS is equal to six. This puts a large burden on the equation $B(\boldsymbol{c})=B_{\min}$, which may give rise to significant differences between the TLS and rTLS models. In particular, this requires that the position of $\boldsymbol{c}$ coincide with an atom.

Eq. 3.20 provides additional insight into the relative contributions of translation and rotation into the B-factors within the TLS model. In the absence of rotation, all B-factors would be the same and equal to $B_{\min}$. Thus the relative contributions of translation and rotation can be evaluated by considering the ratio of the smallest and the largest B-factors, $\kappa=B_{\min}/B_{\max}$. If, for a given protein, κ is close to one, rotation is unimportant. If it is small, rotational motion dominates. For the proteins considered in this study, this

ratio is typically between 0.1 and 0.3 suggesting that the contribution from rotation is always considerable. This is consistent with the recent study by Zheng[173] which also indicates the importance of rotation for accurate representation of B-factors.

C. Extended TLS model (eTLS)

This model recognizes the fact that many proteins have flexible ends that result in large B-factors. Particularly, the 20 protein structures chosen by Song and Jernigan to evaluate their vGNM[155] all share this feature, making a pure TLS description inadequate. To account for the floppiness of the protein extremities, we assume that the bulk of the chain moves as a rigid body while the chain ends are flexible polymers (Fig. 3.1). In the following, this model is referred to as eTLS.

Let $\boldsymbol{p}$ be the position of the last atom treated as part of the rigid body and $\boldsymbol{x}$ the position of an atom belonging to a flexible segment. To calculate

$$B(\boldsymbol{x}) = \left\langle \left(\boldsymbol{x} - \langle \boldsymbol{x} \rangle\right) \cdot \left(\boldsymbol{x} - \langle \boldsymbol{x} \rangle\right) \right\rangle \tag{3.21}$$

we assume that (*i*) $\boldsymbol{p}$ and $\boldsymbol{x} - \boldsymbol{p}$ are statistically independent and (*ii*) The probability distribution of the end-to-end vector $\boldsymbol{x} - \boldsymbol{p}$ of the flexible tail is isotropic so that $\langle \boldsymbol{x} - \boldsymbol{p} \rangle = 0$. These assumptions imply

$$\langle \boldsymbol{x} \rangle = \langle \boldsymbol{p} \rangle + \langle \boldsymbol{x} - \boldsymbol{p} \rangle = \langle \boldsymbol{p} \rangle \tag{3.22}$$

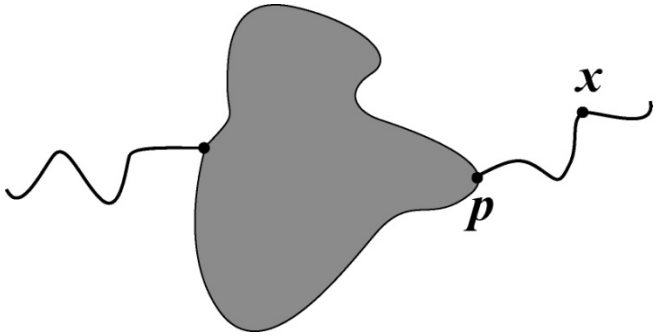


Figure 3.1. Illustration of the eTLS model, which treats the bulk of the protein as a rigid body, except for the chain ends treated as flexible polymers

and

$$B(\boldsymbol{x}) = \left\langle \left(\boldsymbol{x} - \langle \boldsymbol{x} \rangle\right) \cdot \left(\boldsymbol{x} - \langle \boldsymbol{x} \rangle\right) \right\rangle = \left\langle \left(\boldsymbol{x} - \langle \boldsymbol{p} \rangle\right) \cdot \left(\boldsymbol{x} - \langle \boldsymbol{p} \rangle\right) \right\rangle =$$
$$\left\langle \left(\boldsymbol{x} - \boldsymbol{p} + \boldsymbol{p} - \langle \boldsymbol{p} \rangle\right) \cdot \left(\boldsymbol{x} - \boldsymbol{p} + \boldsymbol{p} - \langle \boldsymbol{p} \rangle\right) \right\rangle = \tag{3.23}$$
$$\left\langle \left(\boldsymbol{x} - \boldsymbol{p}\right) \cdot \left(\boldsymbol{x} - \boldsymbol{p}\right) \right\rangle + \left\langle \left(\boldsymbol{p} - \langle \boldsymbol{p} \rangle\right) \cdot \left(\boldsymbol{p} - \langle \boldsymbol{p} \rangle\right) \right\rangle = \left\langle \left(\boldsymbol{p} - \boldsymbol{x}\right)^2 \right\rangle + B(\boldsymbol{p})$$

The mean square distance $\left\langle \left(\boldsymbol{p} - \boldsymbol{x}\right)^2 \right\rangle$ for the flexible tail can be estimated by using a suitable polymer model. In the simplest case, when the polymer chain obeys Gaussian statistics, this quantity is proportional to the number of polymer links k between $\boldsymbol{p}$ and $\boldsymbol{x}$

$$\left\langle \left(\boldsymbol{p} - \boldsymbol{x}\right)^2 \right\rangle = kb^2, \tag{3.24}$$

where b is a constant. At this stage, we can propose a simple extension of the TLS model. Suppose that the part of the chain formed by the atoms from 1 to n_r is treated as rigid

and the remaining part, formed by the atoms $n_r + 1$ to n_p, is treated as flexible. Then the TLS model can be applied to the set of B-factors corresponding to the first n_r atoms. The B-factors for the remaining atoms are fitted with the equation

$$B_i = B_{n_r} + b^2(i - n_r) \qquad i > n_r \tag{3.25}$$

This prescription is easy to modify if one needs to account for two tails.

D. KMSP model

The KMSP model [183] is a particular case of the TLS model. It is obtained by placing the rotation center ***c*** at the center of mass of the protein and assuming that the B-factors are proportional to the square of the distance between the atom and the center of mass. This implies

$$\langle \tilde{\boldsymbol{v}}.\tilde{\boldsymbol{v}} \rangle = 0, \qquad \langle \tilde{\boldsymbol{v}} \times \tilde{\boldsymbol{\omega}} \rangle = 0, \quad \text{and } \langle \boldsymbol{\Omega} \rangle = \langle \tilde{\boldsymbol{\omega}}.\tilde{\boldsymbol{\omega}} \rangle \mathbf{I} \tag{3.26}$$

The validity of these assumptions has not been established.

3.3. Results and discussion

To critically assess the performance of various models of protein dynamics as applied to crystallographic data, we use the experimental data that combines the data sets used in earlier studies. Those data sets include

- The set of 176 proteins used by Eyal and co-workers for evaluating ANM [154],
- The set of 113 proteins used by Kundu and co-workers for evaluating KMSP [183]
- The set of 20 proteins used by Song and Jernigan for evaluating vGNM[155].
- The set of 38 proteins used by Halle [179].

The total number of different proteins involved in these sets is 330, with 17 proteins being involved in more than one set.

To quantify the performance of each model, we use the correlation coefficient [183]

$$\zeta = \frac{\sum_{i=1}^{N}(B_i^{\exp} - B_{avg}^{\exp})(B_i^{com} - B_{avg}^{com})}{\sqrt{\sum_{i=1}^{N}(B_i^{\exp} - B_{avg}^{\exp})^2 \sum_{i=1}^{N}(B_i^{com} - B_{avg}^{com})^2}}, \tag{3.27}$$

where $B^{\exp}$ and B^{com} are the experimental and the computed B-factors and $B_{avg}^{\exp}$ and B_{avg}^{com} are the corresponding averages.

3.3.1. Elastic network models

In this section, we examine two models. The first one is the standard ANM. The second model, referred to as RBM+ANM, is an ENM that accounts for rigid body motion. Specifically, it takes into account the effect of the crystalline environment as prescribed by Eqs. 3.12-13 while its stiffness matrix is constructed as in ANM. Hence for both models the network connectivity is determined by the cut-off radius R and a constant bond stiffness γ is assumed for all pairs of atoms that lie within this cutoff distance.

For the RBM+ANM the crystallographic B-factors are fitted by using the following steps:

1. Choose the cut-off radius R.

2. Construct the intrinsic protein stiffness matrix $\mathbf{K}_{pp}^{i}$ using the bond stiffness $\gamma = 1$.

3. Determine the eigenvectors $\boldsymbol{e}^{\alpha}$ of $\mathbf{K}_{pp}^{i}$.

4. Parametrize the eigenvalues of the correlation matrix as

$$\sigma^{\alpha} = k_B T \begin{cases} \dfrac{1}{\rho^{\alpha}} & \alpha \le 6 \\ \dfrac{1}{\gamma \mu^{\alpha}} & \alpha > 6 \end{cases} \tag{3.28}$$

where ρ^{α} and γ are considered as fitting parameters and μ^{α} are the eigenvalues of $\mathbf{K}_{pp}^{i}$ corresponding to $\gamma = 1$.

5. Determine ρ^{α} and γ from the least square fit of the B-factors corresponding to the covariance matrix

$$C = \sum_{\alpha=1}^{3n_p} \sigma^{\alpha} \boldsymbol{e}^{\alpha} \otimes \boldsymbol{e}^{\alpha} \qquad (3.29)$$

to the experimental data.

Computational results based on the above fitting procedure are demonstrated in Figure 3.2, where we plot the residue number along the abscissa and the B-factors along the ordinate for circularly permuted jellyroll protein (PDB code 1cpn). Both fits correspond to $R = 18\,\mathring{\mathrm{A}}$, which is optimal for ANM [154] but not necessarily for RBM+ANM. The numbers in the legend box are the correlation coefficients. The values $\varsigma = 0.56$ for the ANM and $\varsigma = 0.71$ for the RBM+ANM are typical. Both models recognize but fail to capture accurately a major peak between the 33rd and 49th residues. Further, the ANM predicts two non-existing minor and one major peaks. Failure to capture peaks makes the model unreliable.

An advocate of the ANM may argue that the RBM+ANM relies on eight fitting constants (ρ^{α}, R, and γ) whereas the original model involves only two constants (R

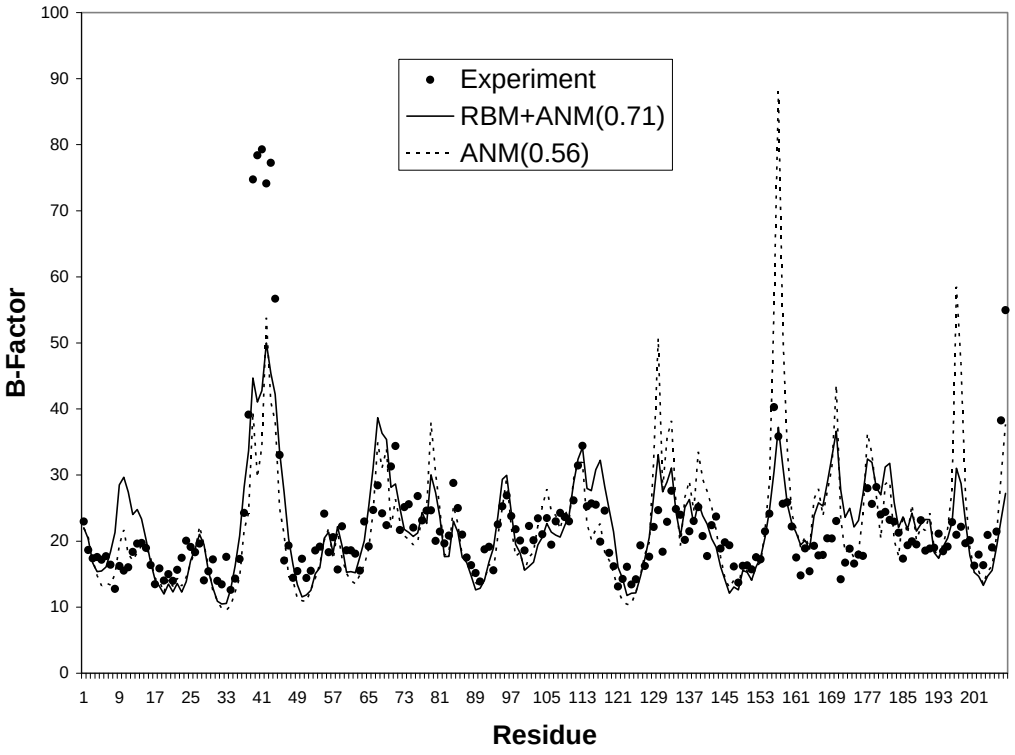


Figure 3.2.B-factors for circularly permuted jellyroll protein (PDB code 1cpn).Comparisons of the ANM and RBM+ANM. The value of the cutoff radius used is R = 18Å [154].

and γ). This argument would totally miss our point made above: There is nothing fundamentally wrong with fitting the stiffness matrix using the constants R and γ but it is fundamentally wrong to neglect rigid body motion. To examine the role of rigid body motion quantitatively, we evaluate the B-factors for the covariance matrix

$$C_{ANM} = \sum_{\alpha=7}^{3n_p} \sigma^{\alpha} \boldsymbol{e}^{\alpha} \otimes \boldsymbol{e}^{\alpha}, \qquad (3.30)$$

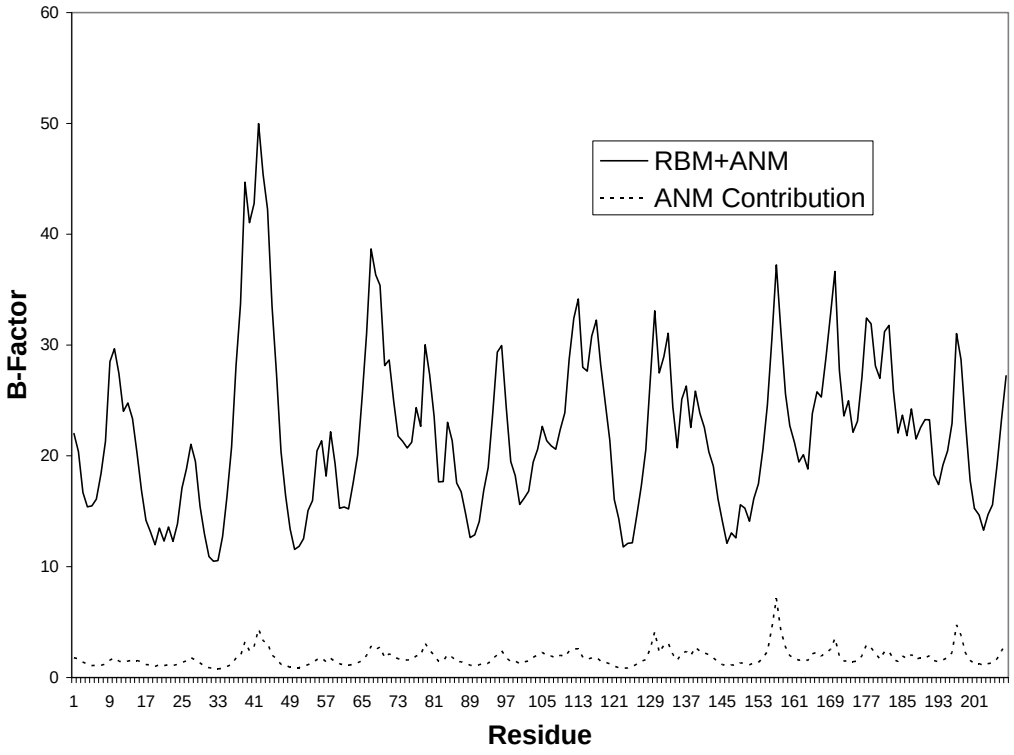


Figure 3.3.B-factors for circularly permuted jellyroll protein (PDB code 1cpn).Evaluation of the ANM contribution for the RBM+ANM. The value of the cutoff radius used is R = 18Å [154].

which represents the ANM contribution of the RBM+ANM. The B-factors corresponding to the matrices **C** and $\mathbf{C}_{ANM}$ are plotted in Figure 3.3. As can be seen from the plot, the B-factors associated with the matrix $\mathbf{C}_{ANM}$ are small, and therefore rigid body motion is the primary contributor to the B-factors.

The smallness of $\mathbf{C}_{ANM}$ makes the calibration of the RBM+ANM unreliable because, in effect, it involves computation of a small difference between two large uncertain numbers. To demonstrate this point consider the data for HIV protease (PDB code 1hhp)

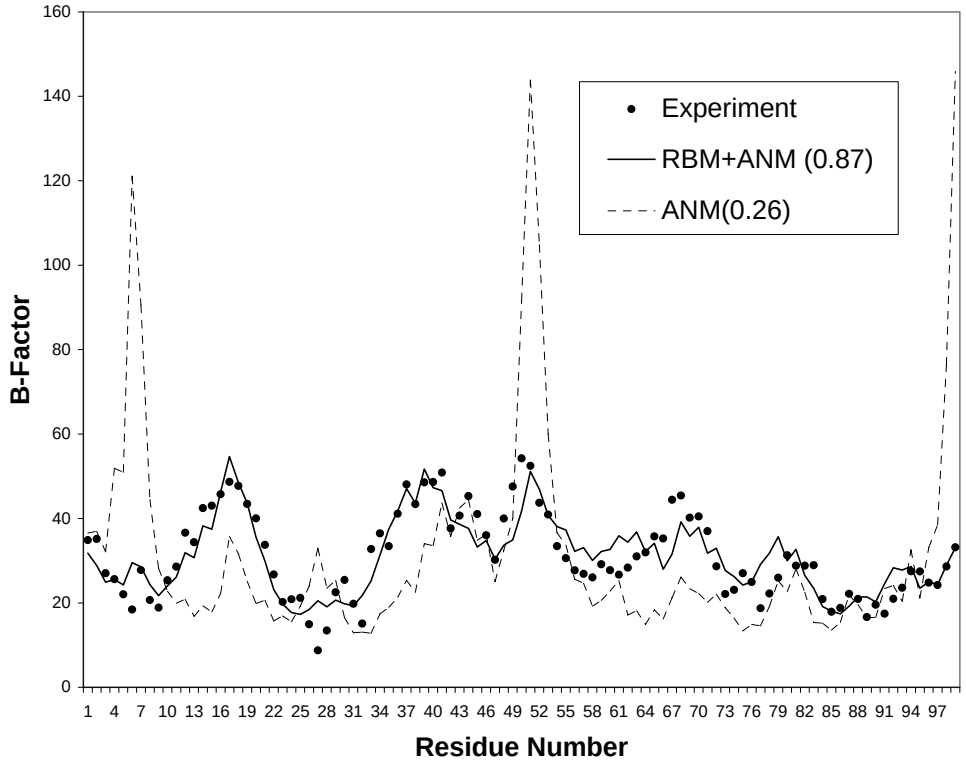


Figure 3.4.B-factors for the HIV protease protein (PDB code 1hhp). Comparisons of the ANM and RBM+ANM, with a cutoff radius *R* = 15Å [154].

shown in Figure 3.4. On the surface, this is a favorable case for RBM+ANM $(\varsigma = 0.87)$ but not for ANM $(\varsigma = 0.26)$. However a closer look exposes that the RBM+ANM predicts an absurdly high (twenty orders of magnitude off) stiffness γ. A possible explanation for this prediction is that the crystallographic B-factors are best represented by rigid body motion plus noise. Since RBM+ANM approximates the noise with the matrix $\mathbf{C}_{ANM}$, which has $3n_p - 6$ positive eigenvalues, it finds that the optimal solution is to force the matrix $\mathbf{C}_{ANM}$ to be zero. As the RBM+ANM fit uses only one adjustable parameter to describe the network properties as opposed to 6 parameters accounting for

rigid-body motion, it is also possible that use of an ANM with more adjustable parameters within the RBM+ANM framework could lead to a better estimate of the network parameters.

In summary, we believe that both ANM and RBM+ANM – if parametrized to fit crystallographic B-factors – are unreliable and that their quantitative use should be treated with suspicion.

3.3.2. TLS models

Here we present case studies featuring the TLS, rTLS, eTLS, and KMSP models. Our goal is twofold. First, we show that *a priori* assumptions leading to simplified expressions for rigid body motion may lead to significant errors. Second, we compare the performance of the TLS and eTLS models and show that the latter is a better choice only when the protein has flexible ends.

TLS versus rTLS and KMSP models. The rTLS and KMSP models are particular cases of the TLS model. As far as the number of fitting parameters is concerned, the differences among the models are summarized in Table 3.1.

Results for myohemerythrin (PDB code 2mhr) are demonstrated in Figure 3.5. The numbers in the legend box are the correlation coefficients, which are typical for these models. It should be noted that the center of mass of the protein postulated by KMSP to be the center of rotation ($\boldsymbol{c}_{KMSP}$) is close to the rotation center $\boldsymbol{c}_{TLS}$ calculated for TLS by

Model	Scalar	Vector	Tensor
TLS	1	1	1
rTLS	0	0	1
KMSP	0	0	0

Table 3.1. Optimality of fitted parameters in the TLS-type models: 1 denotes an optimal choice and 0 a non-optimal choice. Optimality is defined with respect to the least square fit restricted to positive definite tensor-valued fitting parameters.

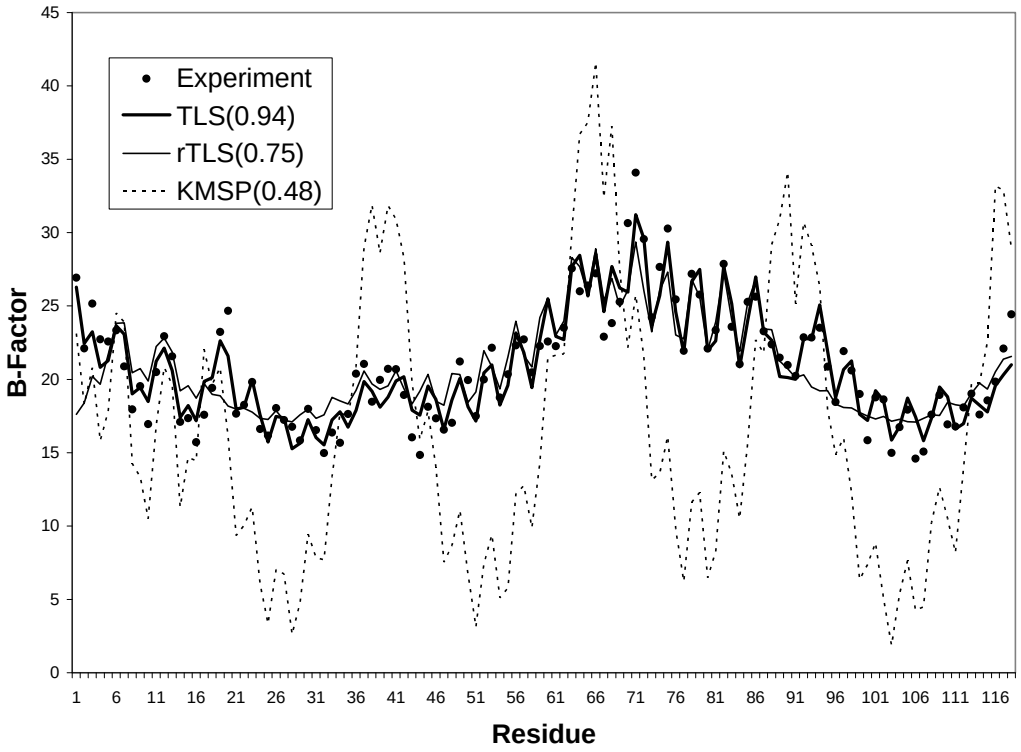


Figure 3.5.B-factors for myohemerythrin (PDB code 2mhr).Comparisons of the TLS, rTLS, and KMSP models.

using Eq. 3.19, $|\boldsymbol{c}_{KMSP} - \boldsymbol{c}_{TLS}| = 2.8\,\text{Å}$, while $|\boldsymbol{c}_{rTLS} - \boldsymbol{c}_{TLS}| = 8.0\,\text{Å}$. Hence the primary source of the poor performance of the KMSP model should be the poor approximation of the second rank tensor rather than a poor choice of the center of rotation.

TLS versus eTLS models. For the eTLS model, we have treated the first and the last three residues as forming flexible tails. We did not attempt to optimize the number of residues to be included in the tails. TLS and eTLS yield significantly different predictions only when flexible tails exist, whose behavior can be captured by eTLS but not by TLS. For example, for peroxidase (PDB code 1aru) the two models have essentially the same correlation coefficient (Fig. 3.6).

In contrast, for myoglobin (PDB code 1abs), the eTLS model significantly outperforms TLS (Fig. 3.7). This is also the case for all of the 20 structures that were used by Song and Jernigan [155] to validate their vGNM.

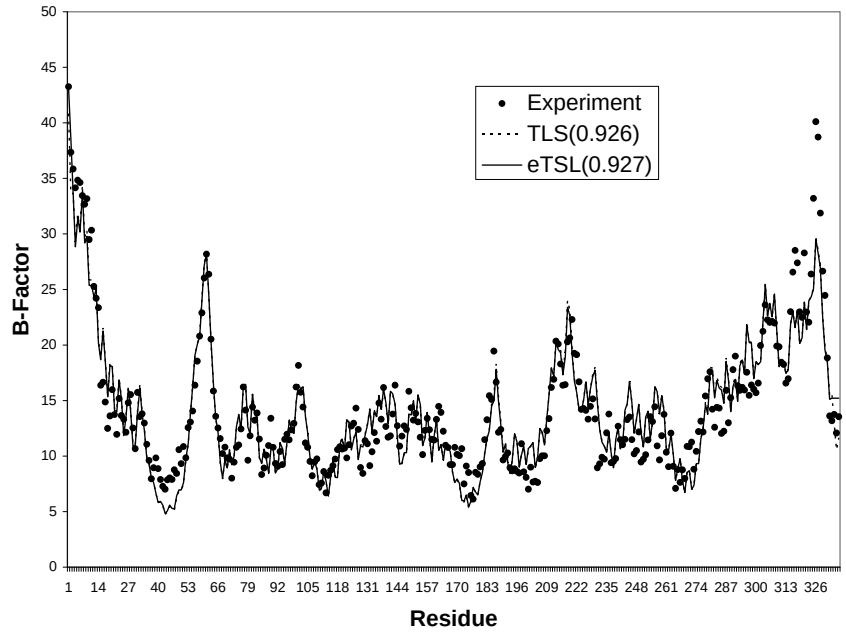


Figure 3.6.B-factors for peroxidase (PDB code 1aru).Comparison of the TLS and eTLS models.

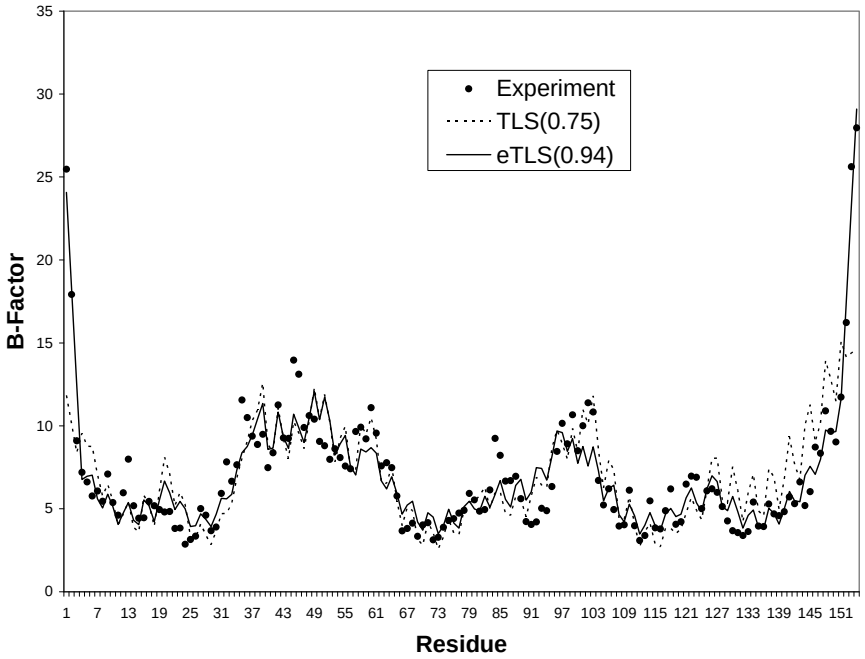


Figure 3.7. B-factors for myoglobin (PDB code 1abs). Comparison of the TLS and eTLS models.

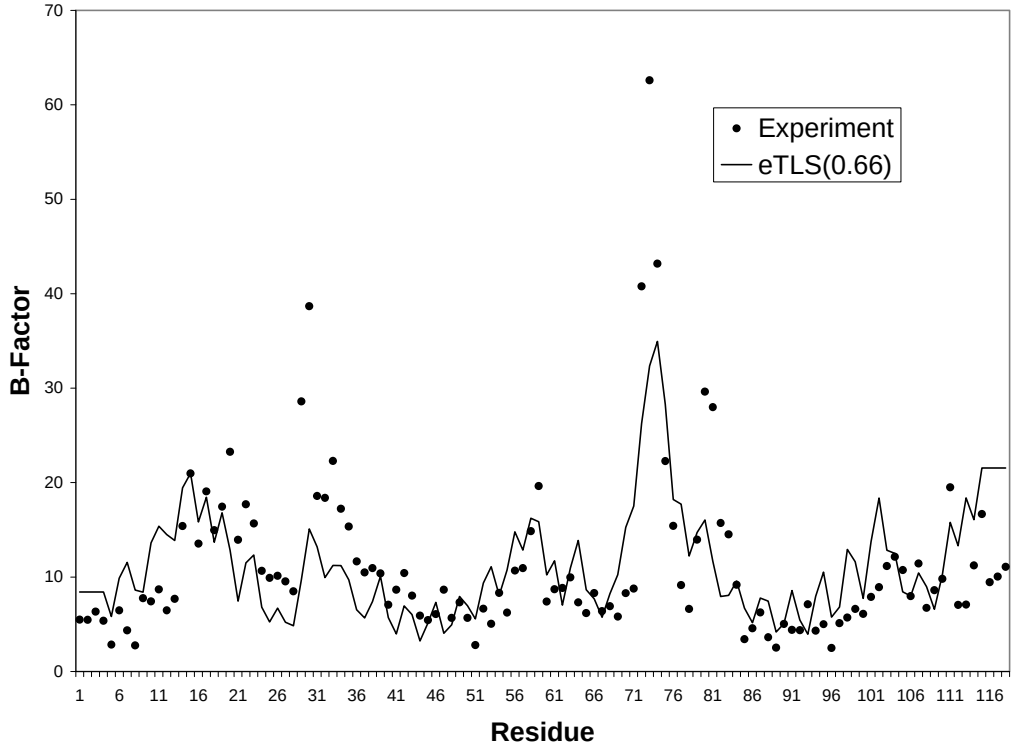


Figure 3.8.B-factors for phospholipase A2 (PDB code 1poa). The eTLS model is not able to capture a peak.

A small subset of proteins cannot be described well by the eTLS model. For example, for phospholipase A2 (PDB code 1poa), $\varsigma = 0.58$. It is apparent from Figure 3.8 that the poor correlation is due to peaks that eTLS fails to capture. In general, disagreement between eTLS and the experimental B-factors may indicate a large-amplitude conformational rearrangement. In particular, proteins that contain more than one subunit often exhibit such motion and, in principle, can be modeled by allowing each subunit to have its own TLS motion [190].

	Eyal et al set[154]	Kundu et al set[183]	Song & Jernigan set [155]	Halle set[179]
ANM	0.55	0.54	0.50	0.50
GNM	0.59	0.59	0.57	0.54
rTLS	0.70	0.68	0.59	0.71
TLS	0.78	0.80	0.68	0.79
vGNM	–	0.81	0.80	–
eTLS	0.82	0.85	0.88	0.84

Table 3.2.Comparisons of ANM, GNM, rTLS, TLS, vGNM and eTLS for four data sets.

3.3.3. Round-Robin comparisons

So far we have presented case studies on selected proteins. Now we use the combined set for 330 proteins to compare the performance of different models. Computational results are summarized in Table 3.2 where we present the average correlation coefficients for each of the four sets.

The comparisons do not involve the RBM+ANM and KMSP models. RBM+ANM is not included because we do not believe that in its current form it is reliable, specially its prediction for the bond stiffness. The KMSP model is not included because it involves unnecessary assumptions that lead to low correlation coefficients.

It is evident from Table 3.2 that the ANM is the least competitive model. Despite its intrinsic limitation of not being able to describe vector-valued displacements with scalar-valued degrees of freedom, we included GNM in the comparisons. Although GNM outperforms ANM, it is outperformed by all models that consider rigid body motion. Both rTLS and TLS are characterized by reasonable correlation coefficients, but predictably, these models are outperformed by eTLS.

Interestingly, the performance of TLS is comparable with that of vGNM for the more diverse data set of Kundu et al[183]. However vGNM outperforms TLS for the 20 myoglobin structures from ref. [155]. All these structures share the feature that the atoms close to the ends of the chain exhibit large fluctuations (cf. Fig. 3.7), suggesting floppy chain ends. It is likely that the GNM part of vGNM captures these fluctuations leading to a better correlation as compared to models that treat proteins as rigid bodies. When the flexibility of the chain ends is accounted for by using eTLS, the resulting average correlation coefficient for these structures becomes much higher (0.88) than that provided by any other model.

Comparisons of vGNM versus eTLS favor the latter model for reasons other than higher correlation coefficients. First, vGNM incorporates a very poor representation of rigid body motion, whereas the TLS model clearly indicates that rigid body motion is the main contributor to the B-factors. Second, the idea of enriching the correlation matrix corresponding to rigid body motion with a low-rank matrix whose basis is constructed using the eigenvectors of GNM is sensible but it would be more meaningful to use the

eigenvectors of an ANM rather than a GNM. Finally, vGNM typically uses 21 fitting parameters whereas eTLS relies on 13, each of which having a clear physical meaning. The correlation coefficients for vGNM are taken directly from the original paper [155] and we did not apply the model to the first and fourth data sets.

3.4. Conclusions and outlook.

The development of elastic network models has been largely motivated by their ability to capture slow and often biologically important internal modes of proteins at a modest computational expense. As they offer an appealing alternative to computationally prohibitive all-atom simulations, there is a considerable interest in using ENMs to answer more quantitative questions concerning protein dynamics. Network models are particularly encouraging in the context of quantifying mechanical properties of proteins and their assemblies, where they provide a natural framework similar to the truss model in structural mechanics. To address quantitative issues it is crucial to develop reliable calibration procedures for ENM. In this work, we have critically evaluated one widespread calibration approach based on fitting the crystallographic B-factors. Our main findings are as follows:

- Crystallographic B-factors for most proteins that have previously been employed to validate and/or parametrize elastic network models and Gaussian network models are dominated by rigid body motion. As a result, such experimental data are poorly suited for the calibration of models describing internal protein motions.

- The eTLS model, which considers the protein as a rigid body with flexible ends, outperforms all other simple protein-databank based models for protein dynamics in crystals and yields an average correlation coefficient above 0.8 for a set of 330 proteins used in previous studies by other researchers.

- The calibration procedures commonly used for ANM and GNM are fundamentally flawed because they fail to recognize the importance of rigid body motion and crystalline constraints.

We emphasize that in no way does our work question the usefulness of simple network models. Rather we exposed flaws in the existing calibration procedures and question the usefulness of experimental B-factors for this purpose.

It is natural to ask whether it is possible to extract information about internal protein dynamics based on the small residual difference between the B-factors and the TLS or eTLS fits. In principle, it can be accomplished through normal mode refinement [185, 186]. In practice, however, this approach requires an initial normal mode calculation that is based on a realistic molecular dynamics force field. Considering other limitations of the reported crystallographic B-factors mentioned in the beginning, it may be better to forego using the B-factors altogether and to rely on molecular dynamics to calibrate ENMs[176, 177, 191].

Finally, if the crystallographic B-factors are dominated by rigid body motion, why are they correlated at all with models that neglect rigid body motion altogether, such as ANM or GNM? One possible explanation is the correlation between the B-factor and the

proximity of the atom to the surface[157], which is captured by all of these models even if not for the right physical reason. In TLS, the further the atom is from the rotation center the larger the displacement. Thus peripheral atoms should have larger B-factors, which is qualitatively the trend also observed in ANM/GNM[157, 179].

Chapter Four

Polymer mediated interaction between two planar surfaces

4.1. Introduction

There is a considerable interest in studying the interaction between polymers and solid surfaces as well as polymer-mediated interactions between surfaces[192, 193]. Interactions of this type are commonly encountered in polymer science[194], biophysics[195] (e.g., cell adhesion) and study of colloids[196, 197]. An important example of this interaction is encountered in colloidal systems, where the polymers covering the surfaces of colloidal particles cause repulsion between the particles and stabilize the colloids against van der walls attraction[198].

Polymers confined between two surfaces can be free or attached to the surfaces, depending on the details of their interaction with the surfaces. In the case where they attach to the surfaces, polymers can form a tail, a loop, a bridge, or a train[199] (see Figure 4.1). Tail is the case where only one end of the chain is attached to a surface. Loop is created when the two ends of the chains attach to the same surface, and bridge is made when the two ends attach to different surfaces. Train is the case where the polymers are in direct contact with the surface.

The statistical mechanics of confined polymers has been studied extensively, both theoretically and experimentally(see [200] and references therein). On a theoretical side, while there have been many numerical studies involving Monte Carlo simulations

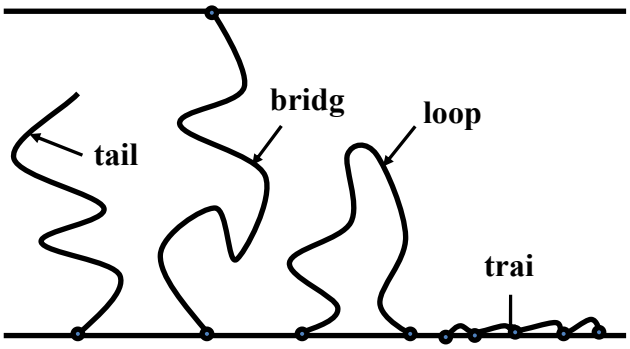


Figure 4.1.Schematics of the formation of a tail, a bridge, a loop, and a train between two surfaces.

[201, 202], the transition matrix method[194, 203], and other approaches, analytical results have been considerably more sparse[198, 204].

Dolan and Edwards[198] derived, analytically, the free energy of a tail confined between two parallel plane surfaces, by modeling the polymer as an ideal chain. Later, they have also incorporated excluded volume effects in their model[205]. Bhatia and Russel[204]have studied, analytically, the problem of a bridge between two surfaces, which can be either planar or spherical.

In this chapter, we present analytical results for the problem of a polymer between two planar surfaces, assuming that this polymer is an ideal chain. Specifically, we consider a thin polymer layer confined between two parallel plates that are a distance h apart, where h is less than the contour length of the polymer chains. We change the distance slowly and calculate the external force that needs to be applied to the surfaces to hold the system in equilibrium. Thus a force-distance diagram is calculated for the system.

We further consider two different models for the confined polymer. In model A, the polymer layer consists of a large number of individual chains, each having the same contour length (monodisperse polymer). Every chain makes a bridge between two plates by attaching each of its ends to a different plate.

In model B, a very long polymer chain is confined between two plates. After reaching equilibrium at a specific distance, any bead that contacts the plates makes a permanent bond to them at that point. This results in segmentation of the polymer chains into tails, loops and bridges. Here, as we only have one long chain, there are, at most, two tails after segmentation. In contrast, the number of loops and bridges can be large as the chain is long. Hence, we can neglect the effect of the tails. In this study, we focus on the role of the bridges. In effect, Model B describes a polydisperse polymer in which the distribution of its length is governed by Gaussian statistics, Allegra et al (2003) have studied similar models numerically by applying transition-matrix methods to two lattice models of the polymer.

In both model A and B, we assume that the polymer chains are in the mushroom regime[200] and neglect excluded volume interactions both among different chains and within each individual chain.

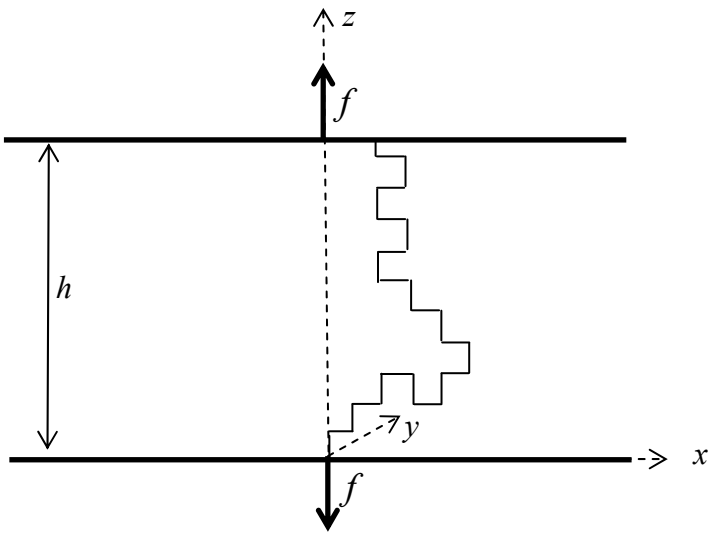


Figure 4.2. A polymer chain bridging two plates

4.2. Single chain forming a bridge (Model A)

Here, we follow the approach of Dolan and Edwards[198] to obtain the free energy and force-distance relationship for a single polymer making a bridge between two plates. Bhatia and Russel[204] have studied a similar problem and considered both planar and spherical surfaces.

To derive an expression for the free energy of a single confined polymer chain, we assume that the chain is ideal and obeys Gaussian statistics. Hence, we can use polymer Green's functions[87] to describe its statistics. Polymer Green's function $G\left(\begin{matrix}1\\ \boldsymbol{r_0}\end{matrix}\middle|\begin{matrix}N\\ \boldsymbol{r}\end{matrix}\right)$ is defined as the probability that we find the Nth bead at $\boldsymbol{r} = (x, y, z)$when the first bead is located at $\boldsymbol{r_0} = (x_0, y_0, z_0)$. This function is governed by the following differential equation:

$$\left(\frac{\partial}{\partial N} - \frac{b^2}{6}\nabla^2\right) G\left(\begin{matrix}1\\ \boldsymbol{r_0}\end{matrix}\middle|\begin{matrix}N\\ \boldsymbol{r}\end{matrix}\right) = \delta^3(\boldsymbol{r} - \boldsymbol{r_0})\delta(N) \qquad (4.1)$$

where$b$ is the polymer bond length.

Because the chain cannot penetrate the surfaces, the absorbing boundary condition $G\left(\begin{matrix}1\\ \boldsymbol{r_0}\end{matrix}\middle|\begin{matrix}N\\ \boldsymbol{r}\end{matrix}\right) = 0$ must be imposed at the confining surfaces. The effect of this boundary condition is the removal of all the configurations that cross the boundaries.

Let us assume that two plates are normal to z–axis, and the bottom and top plates are located, respectively, at $z = 0$ and $z = h$ (Figure 4.2). Given the planar geometry, the solution of Eq. 4.1 is separable and can be written as

$$G\left(\begin{matrix}1\\ \boldsymbol{r_0}\end{matrix}\middle|\begin{matrix}N\\ \boldsymbol{r}\end{matrix}\right) = g_x\left(\begin{matrix}1\\ x_0\end{matrix}\middle|\begin{matrix}N\\ x\end{matrix}\right) g_y\left(\begin{matrix}1\\ y_0\end{matrix}\middle|\begin{matrix}N\\ y\end{matrix}\right) g_z\left(\begin{matrix}1\\ z_0\end{matrix}\middle|\begin{matrix}N\\ z\end{matrix}\right) \qquad (4.2)$$

whereg_α is the one-dimensional Green's function in direction α; $\alpha = \{x, y, z\}$. Furthermore, the only macroscopic change in the system is in the z direction, and the statistics in the x and$y$ directions are not affected by this change. Hence, we only need to consider the z-component of the probability distribution function$g_z\left(\begin{matrix}1\\ z_0\end{matrix}\middle|\begin{matrix}N\\ z\end{matrix}\right)$. In the rest of the chapter, we drop the subscript z and call this quantity simply $g\left(\begin{matrix}1\\ z_0\end{matrix}\middle|\begin{matrix}N\\ z\end{matrix}\right)$.

Two forms of the solution can be obtained for $g\left(\begin{matrix}1\\ z_0\end{matrix}\middle|\begin{matrix}N\\ z\end{matrix}\right)$. The method of eigenfunction expansion results in the following solution[206]:

$$g\left(\begin{matrix}1\\z_0\end{matrix}\middle|\begin{matrix}N\\z\end{matrix}\right)=\frac{2}{h}\sum_{n=1}^{\infty}Sin\left[\frac{n\pi z}{h}\right]Sin\left[\frac{n\pi z_0}{h}\right]Exp[-\frac{n^2\pi^2Nb^2}{6h^2}] \quad (4.3)$$

The second solution is obtained by the method of images[207]:

$$g\left(\begin{matrix}1\\z_0\end{matrix}\middle|\begin{matrix}N\\z\end{matrix}\right)=$$

$$\sqrt{\frac{3}{2\pi Nb^2}}\sum_{n=-\infty}^{\infty}\left\{Exp\left[-\frac{3}{2\pi Nb^2}(z-z_0-2nh)^2\right]-Exp\left[-\frac{3}{2\pi Nb^2}(z+z_0-2nh)^2\right]\right\} \quad (4.4)$$

The two forms are identical[206]. However, both forms will be used as each one is more convenient than the other in certain limits.

For the case of a bridge, we assume that the first bead is attached to the bottom plate. However, using this point for the position of the first bead results in the trivial solution $g\left(\begin{matrix}1\\0\end{matrix}\middle|\begin{matrix}N\\z\end{matrix}\right)=0$ for Eq. 4.1. This difficulty is readily overcome using the standard trick [198, 204] of placing the first bead at some finite but small distance $z_0=\varepsilon$. Physically, this distance should be comparable with the bond length of the polymer chain. Since we assume that the chain's contour length is much longer than a single bond length, we will eventually take the limit of ε going to zero.

The probability $p_b(h,N)$ (b stands for bridge) that a chain of N monomers, which is confined between two plates at distance h, makes a bridge, can be calculated as

$$p_b(h,N)=\lim_{\varepsilon\to 0} g\left(\begin{matrix}1\\\varepsilon\end{matrix}\middle|\begin{matrix}N\\h-\varepsilon\end{matrix}\right) \quad (4.5)$$

Using Eq. 4.3, this gives

$$p_b(h,N)=\lim_{\varepsilon\to 0}\frac{-2}{h}\sum_{n=1}^{\infty}Cos[n\pi]\left(Sin\left[\frac{n\pi\varepsilon}{h}\right]\right)^2Exp[-\frac{n^2\pi^2Nb^2}{6h^2}] \quad (4.6)$$

Expanding this to the second order in ε, the Eq. 4.6 reduces to the following

$$p_b(h,N) = \frac{-2\pi^2\varepsilon^2}{h^3}\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp[-\frac{n^2\pi^2 Nb^2}{6h^2}] \tag{4.7}$$

The free energy of the bridge $A_b(h,N)$ is calculated from this probability as

$$A_b(h,N) = -k_B T \ln p_b(h,N) + C = k_B T\left(3\ln h - ln\left[\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp\left[-\frac{n^2\pi^2 Nb^2}{6h^2}\right]\right]\right) + C_1 \tag{4.8}$$

wherek_B is Boltzmann's constant, T is the temperature and C and C_1 are constants independent of h. The external force $f_b(h,N)$, which is required to hold the system in equilibrium (see Figure 4.2) is obtained by

$$f_b(h,N) = \frac{dA_b(h,N)}{dh} = -k_B T \frac{1}{p_b(h,N)}\frac{dp_b(h,N)}{dh} \tag{4.9}$$

This yields the following equation for $f_b(h,N)$,

$$f_b(h,N) = \frac{k_B T}{h}\left\{3 - \frac{\pi^2 Nb^2}{3h^2}\frac{\sum_{n=1}^{\infty} n^4 Cos[n\pi] Exp[-\frac{n^2\pi^2 Nb^2}{6h^2}]}{\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp[-\frac{n^2\pi^2 Nb^2}{6h^2}]}\right\} \tag{4.10}$$

Let us define the non-dimensional force and distance by $\hat{f} = \frac{f\,h}{k_B T}$ and $\xi = \frac{h}{b\sqrt{N}}$, respectively. Then, Eq. 4.10 reduces to

$$\hat{f}_b(\xi) = 3 - \frac{\pi^2}{3\xi^2}\frac{\sum_{n=1}^{\infty} n^4 Cos[n\pi] Exp[-\frac{n^2\pi^2}{6\xi^2}]}{\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp[-\frac{n^2\pi^2}{6\xi^2}]} \tag{4.11a}$$

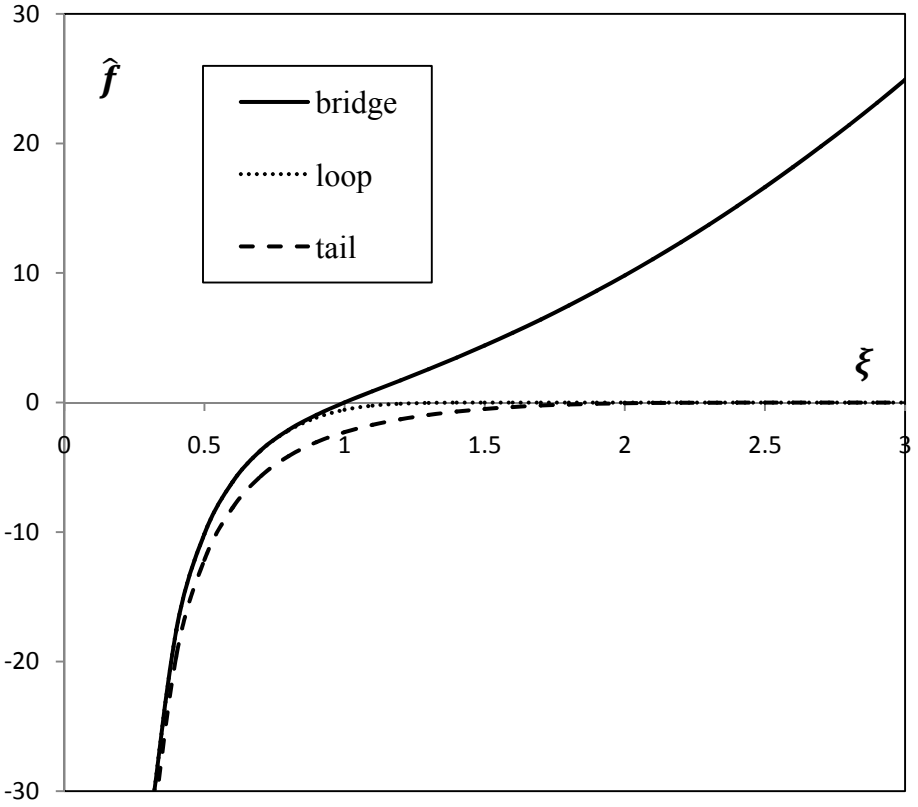


Figure 4.3 Force-distance curves for three cases of bridge, loop and tail.

This equation can be represented in terms of elliptic theta function as

$$\hat{f}_b(\xi) = 3 - \frac{\pi^2}{3\xi^2}\left(1 + Exp[-\frac{\pi^2}{6\xi^2}]\frac{\vartheta_4''(0,Exp[-\frac{\pi^2}{6\xi^2}])}{\vartheta_4'(0,Exp[-\frac{\pi^2}{6\xi^2}])}\right) \qquad (4.11b)$$

whereϑ_4 is the elliptic theta function of the fourth kind and prime denotes differentiation with respect to the second argument, e.g. $\vartheta_4'(0,u) = \frac{d\vartheta_4(0,u)}{du}$.

Figure 4.3 shows the normalized force $\hat{f}_b$ versus the normalized distance ξ. As is seen from the plot, the equilibrium distance between two plates corresponds to $\xi = 1$,

where the force become zero. For $\xi > 1$, two plates attract each other, while for $\xi < 1$, they repel each other. For comparison, the force in the tails and loops are also plotted in Figure 4.3. The force in the case of loops and tail are always repulsive and approaches to zero at long distances. We see from the plot that the repulsive part of the curve for the bridge is very close to that of a loop.

Asymptotic behavior

Here, we study two asymptotic limits for the bridges. If the distance between two plates is very small, i.e. $\xi \ll 1$, then the first term of the sums in Eq. 4.10 dominates. As a result, the equation simplifies to

$$\hat{f}_b(\xi) \cong 3 - \frac{\pi^2}{3\xi^2} \cong -\frac{\pi^2}{3\xi^2} \tag{4.12a}$$

This can be written as

$$f_b(h, N) = -k_B T \frac{\pi^2 N b^2}{3h^3} \tag{4.12b}$$

The same asymptotic limit as Eq. 4.12b is obtained for the case of a loop and a tail, showing that, at small distance, the behavior of all of them is the same.

At large distances, it is more convenient to use the second representation for Green's function given by Eq. 4.4. Substituting this into Eq. 4.5 and taking the limit of small ε, we obtain

$$p_b(h, N) = -\sqrt{\frac{3}{2\pi N b^2}} \frac{12\varepsilon^2}{Nb^2} \sum_{\substack{n=1 \\ n\ odd}}^{\infty} Exp[-\frac{3h^2}{2Nb^2} n^2] \tag{4.13}$$

Using the same steps as the other solution, we obtain the normalized force as

$$\hat{f}_b(\xi) = 3\xi^2 \frac{\sum_{\substack{n=1 \\ n\,odd}}^{\infty} n^2 Exp[-\frac{3n^2\xi^2}{2}]}{\sum_{\substack{n=1 \\ n\,odd}}^{\infty} Exp[-\frac{3n^2\xi^2}{2}]} \tag{4.14}$$

Now, for $\xi \gg 1$, the first term of the sums in Eq. 4.14 dominates, and we get

$$\hat{f}_b(\xi) \cong 3\xi^2 \tag{4.15a}$$

This can be rewritten as

$$f_b(h, N) \cong k_B T \frac{3h}{Nb^2} \tag{4.15b}$$

This means that for large values of h, the force is linearly proportional to the distance.

4.3. Model B

In this model, we assume that the contour length of a bridge is not constant but rather is in itself a randomly distributed variable. More specifically, we will consider the case where the configuration of each bridge is effectively an unbiased random walk between the two plates held some predefined distance h_0 apart. Such a scenario may, for example, arise if the first surface, with an array of polymer chains tethered to it, is quickly brought to proximity with the second surface, provided that the chain monomers can irreversibly attach to the second surface.

If the two plates are held at a fixed distance h_0, then the probability $p_{h_0}(N)$ that a chain bridging those two plates has N beads is proportional to the flux[208, 209]$J(z,N)$ at $z = h_0$, defined as

$$J(z,N) = \frac{\partial g\left(\frac{1}{\varepsilon}\middle|\frac{N}{z}\right)}{\partial z} \tag{4.16}$$

Hence, $p_{h_0}(N)$ can be calculated as

$$p_{h_0}(N) = \lim_{\varepsilon\to 0} \frac{J(h_0,N)}{\int_{N_0}^{\infty} J(h_0,N)dN} \tag{4.17}$$

where the denominator is a normalization factor. Since the contour length of a chain forming a bridge cannot be shorter than the distance between the plates, the lower integration limit is given by $N_0 = \frac{h_0}{b}$.

Replacing Eq. 4.4 in Eq. 4.16 yields

$$J(z,N) = \frac{2\pi}{h^2}\sum_{n=1}^{\infty} nCos\left[\frac{n\pi z}{h}\right] Sin\left[\frac{n\pi\epsilon}{h}\right] Exp[-\frac{n^2\pi^2 Nb^2}{6h^2}] \tag{4.18}$$

As a result, the equation for the probability becomes

$$p_{h_0}(N) = \lim_{\varepsilon\to 0}\left(-\frac{\pi^2 b^2}{6h_0^2}\frac{\sum_{n=1}^{\infty} nCos[n\pi]Sin\left[\frac{n\pi\epsilon}{h_0}\right]Exp[-\frac{n^2\pi^2 Nb^2}{6h_0^2}]}{\sum_{n=1}^{\infty}\frac{1}{n}Cos[n\pi]Sin\left[\frac{n\pi\epsilon}{h_0}\right]Exp[-\frac{n^2\pi^2 N_0 b^2}{6h_0^2}]}\right) \tag{4.19}$$

Taking the limit, we obtain

$$p_{h_0}(N) = -\frac{\pi^2 b^2}{3h_0^2}\frac{\sum_{n=1}^{\infty} n^2 Cos[n\pi]Exp[-\frac{n^2\pi^2 Nb^2}{6h_0^2}]}{\sum_{n=1}^{\infty} Cos[n\pi]Exp[-\frac{n^2\pi^2 b}{6h_0}]} \tag{4.20}$$

This can further simplified to

$$p_{h_0}(N) = -\frac{\pi^2 b^2}{3h_0^2}\frac{\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp[-\frac{n^2\pi^2 Nb^2}{6h_0^2}]}{1-\vartheta_4(0,-\frac{\pi^2 b}{6h_0})} \tag{4.21}$$

Here, $\vartheta_4(0,-\frac{\pi^2 b}{6h_0})$ approaches zero as $\frac{b}{h_0}$ goes to zero. Assuming that the distance between the plates is much longer than the polymer bond length, $\frac{b}{h_0} \ll 1$, the denominator becomes equal to 1 and the equation reduces to:

$$p_{h_0}(N) = -\pi^2 \frac{b^2}{3h_0^2}\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp\left[-\frac{n^2\pi^2 Nb^2}{6h_0^2}\right] \tag{4.22a}$$

or

$$p_{h_0}(N) = \pi^2 \frac{b^2}{6h_0^2} Exp\left[-\frac{\pi^2 Nb^2}{6h_0^2}\right] \vartheta_4'(0,-\frac{\pi^2 Nb^2}{6h_0^2}) \tag{4.22b}$$

We assume that, as the distance h between the plates is varied, the chains cannot detach from or reattach to the surfaces; That is, the distribution of the chain lengths is fixed and remains independent of the distance h. It, however, depends on the initial distance h_0, at which the ensemble was created. So, the average free energy per single chain is calculated as

$$\langle A_b(h, h_0, N_0)\rangle = \int_{N_0}^{\infty} A_b(h, N) p_{h_0}(N) dN \tag{4.23}$$

Substituting Eqs. 4.8 and 4.22a in Eq. 4.23, yields

$$\frac{\langle A_b(h,h_0,N_0)\rangle}{k_BT} = 3\ln h + C_1 +$$

$$\pi^2\frac{b^2}{3h_0^2}\int_{N_0}^{\infty} ln\left[\sum_{n=1}^{\infty} n^2 Cos[n\pi]Exp\left[-\frac{n^2\pi^2Nb^2}{6h^2}\right]\right]\left(\sum_{n=1}^{\infty} n^2 Cos[n\pi]Exp\left[-\frac{n^2\pi^2Nb^2}{6h_0^2}\right]\right)dN$$

(4.24)

Now, let us define normalized length, distance, and free energy as $\tau = \frac{Nb^2}{h_0^2}$, $\hat{h} = \frac{h}{h_0}$and $\hat{A} = \frac{A}{k_BT}$, respectively. We can rewrite the Eq. 4.24 as

$$\langle \hat{A}_b(\hat{h},\tau_0)\rangle = 3\ln\hat{h} + C_2 +$$

$$\frac{\pi^2}{3}\int_{\tau_0}^{\infty} ln\left[\sum_{n=1}^{\infty} n^2 Cos[n\pi]Exp\left[-\frac{n^2\pi^2\tau}{6\hat{h}^2}\right]\right]\left(\sum_{n=1}^{\infty} n^2 Cos[n\pi]Exp\left[-\frac{n^2\pi^2\tau}{6}\right]\right)d\tau \quad (4.25)$$

where $\tau_0 = \frac{b}{h_0}$. The average normalized force $\langle \hat{f}_b(\hat{h},\tau_0)\rangle = \frac{\langle f(\hat{h},\tau_0)\rangle h_0}{k_BT}$ per single chain is then calculated as

$$\langle \hat{f}_b(\hat{h},\tau_0)\rangle = \frac{3}{\hat{h}} + \frac{\pi^4}{9\,\hat{h}^3}\int_{\tau_0}^{\infty}\frac{\sum_{n=1}^{\infty} n^4 Cos[n\pi]Exp\left[-\frac{n^2\pi^2\tau}{6\hat{h}^2}\right]}{\sum_{n=1}^{\infty} n^2 Cos[n\pi]Exp\left[-\frac{n^2\pi^2\tau}{6\hat{h}^2}\right]}\left(\tau\sum_{n=1}^{\infty} n^2 Cos[n\pi]Exp\left[-\frac{n^2\pi^2\tau}{6}\right]\right)d\tau$$

(4.26)

This integral can be written in terms of elliptic theta function as

$$\langle \hat{f}_b(\hat{h},\tau_0)\rangle =$$

$$\frac{3}{\hat{h}} + \frac{\pi^4}{18\,\hat{h}^3}\int_{\tau_0}^{\infty}\tau Exp\left[-\frac{\pi^2\tau}{6}\right]\vartheta_4'(0,Exp[-\frac{\pi^2\tau}{6}])\left(1 + Exp\left[-\frac{\pi^2\tau}{6\hat{h}^2}\right]\frac{\vartheta_4''(0,Exp[-\frac{\pi^2\tau}{6\hat{h}^2}])}{\vartheta_4'(0,Exp[-\frac{\pi^2\tau}{6\hat{h}^2}])}\right)d\tau \quad (4.27)$$

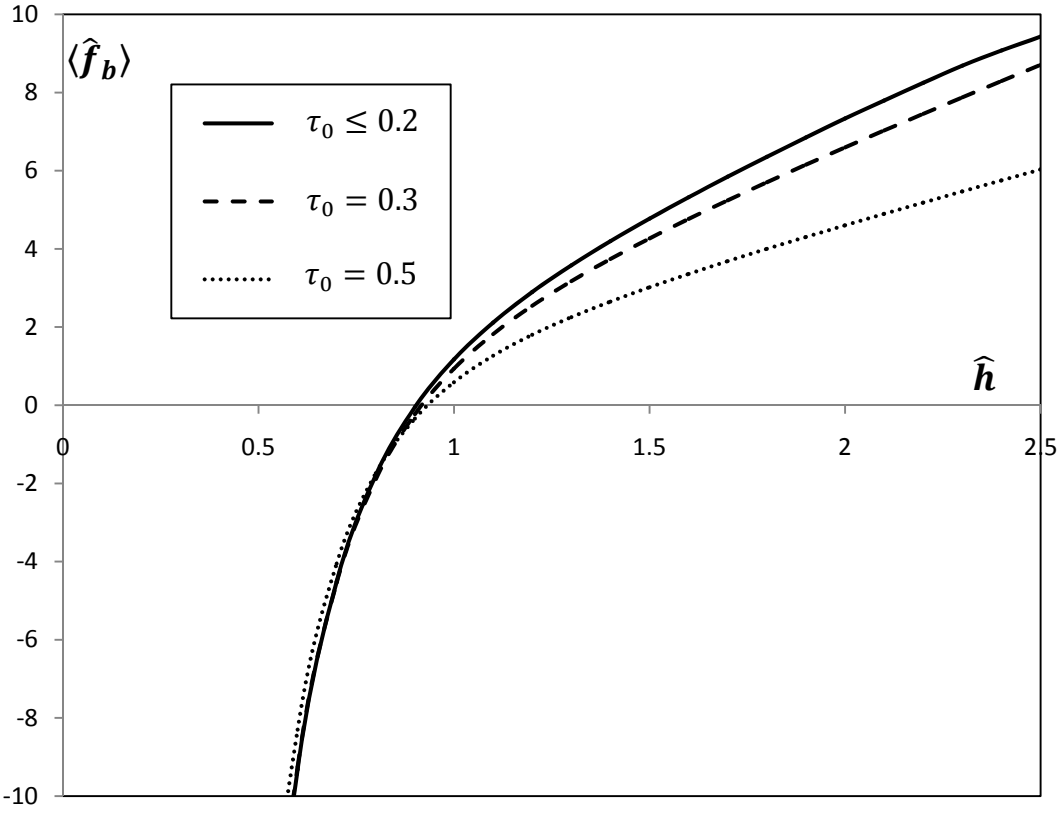


Figure 4.4 Force-distance curves for model B for different values of τ_0.

The plot in Figure 4.4 shows $\langle \hat{f}_b \rangle$ versus $\hat{h}$, for different values of τ_0. As can be seen from the plot, for $\tau_0 = \frac{b}{h_0} \leq 0.2$, the result is virtually independent of τ_0. The average force becomes zero at about $\hat{h} = 0.92$. At small h, it becomes repulsive and at large h, it is attractive. Specifically, at the initial distance $\hat{h} = 1$, there is an attraction between the plates and the normalized force is equal to 1.

Asymptotic behavior

Let us consider the asymptotic behavior at small and large values of $\hat{h}$. For $\hat{h} \ll 1$, the fraction in the integral kernel in Eq. 4.26 becomes equal to 1. Hence, the equation is simplified as

$$\langle \hat{f}_b(\hat{h}, \tau_0) \rangle = \frac{3}{\hat{h}} + \frac{\pi^4}{9\,\hat{h}^3} \int_{\tau_0}^{\infty} \tau \left(\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp\left[-\frac{n^2\pi^2\tau}{6} \right] \right) d\tau \qquad (4.28)$$

By taking the integral, Eq. 4.28 reduces to

$$\langle \hat{f}_b(\hat{h}, \tau_0) \rangle = \frac{3}{\hat{h}} + \frac{2}{3\,\hat{h}^3} \sum_{n=1}^{\infty} \frac{(6+\pi^2 n^2 \tau_0)}{n^2} Cos[n\pi] Exp\left[-\frac{n^2\pi^2\tau_0}{6} \right] \qquad (4.29)$$

For $\tau_0 \ll 1$, the sum in Eq. 4.29 is equal to $-\frac{\pi^2}{2}$. As a result, the average force in this limit becomes

$$\langle \hat{f}_b(\hat{h}) \rangle \cong \frac{3}{\hat{h}} - \frac{\pi^2}{3\,\hat{h}^3} \cong -\frac{\pi^2}{3\,\hat{h}^3} \qquad (4.30a)$$

or

$$\langle f_b(h, h_0) \rangle \cong -k_B T \frac{\pi^2 h_0^2}{3\,h^3} \qquad (4.30b)$$

To find the asymptotic behavior for $\hat{h} \gg 1$, we use the second solution to Green's function for the free energy in a single bridge. However, we still use the Eq. 4.22a for the chain length distribution. As a result, we obtain:

$$\frac{\langle A_b(h,h_0,N_0)\rangle}{k_BT} =$$

$$\pi^2 \frac{b^2}{3h_0^2} \int_{N_0}^{\infty} ln\left[\sum_{\substack{n=1 \\ n\ odd}}^{\infty} Exp[-\frac{3h^2}{2Nb^2}n^2]\right]\left(\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp\left[-\frac{n^2\pi^2 Nb^2}{6h_0^2}\right]\right) dN \quad (4.31)$$

which can be represented in a non-dimensional form as

$$\langle \hat{A}_b(\hat{h},\tau_0)\rangle =$$

$$\frac{\pi^2}{3} \int_{\tau_0}^{\infty} ln\left[\sum_{\substack{n=1 \\ n\ odd}}^{\infty} Exp[-\frac{3\hat{h}^2}{2\tau}n^2]\right]\left(\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp\left[-\frac{n^2\pi^2\tau}{6}\right]\right) d\tau (4.32)$$

The force is then calculated as

$$\langle \hat{f}_b(\hat{h},\tau_0)\rangle = -\pi^2 \hat{h} \int_{\tau_0}^{\infty} \frac{\sum_{\substack{n=1 \\ n\ odd}}^{\infty} n^2 Exp[-\frac{3\hat{h}^2}{2\tau}n^2]}{\sum_{\substack{n=1 \\ n\ odd}}^{\infty} Exp[-\frac{3\hat{h}^2}{2\tau}n^2]} \left(\frac{1}{\tau}\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp\left[-\frac{n^2\pi^2\tau}{6}\right]\right) d\tau \quad (4.33)$$

Now, for the limit where $\hat{h} \gg 1$, the fraction in Eq. 4.33 becomes equal to 1 and the equation reduces to

$$\langle \hat{f}_b(\hat{h},\tau_0)\rangle = -\pi^2 \hat{h} \int_{\tau_0}^{\infty} \frac{\sum_{n=1}^{\infty} n^2 Cos[n\pi] Exp\left[-\frac{n^2\pi^2\tau}{6}\right]}{\tau} d\tau \quad (4.34)$$

Taking the integral we obtain

$$\langle \hat{f}_b(\hat{h},\tau_0)\rangle = -\pi^2 \hat{h} \sum_{n=1}^{\infty} n^2 Cos[n\pi] \Gamma\left[0, -\frac{n^2\pi^2\tau_0}{6}\right] \quad (4.35)$$

whereΓ is the incomplete Gamma function. For $\tau_0 \ll 1$ the above equation gives

$$\langle \hat{f}_b(\hat{h},\tau_0)\rangle \cong 4.2072\ \hat{h} \quad (4.36a)$$

or

$$\langle f_b(h,h_0)\rangle \cong 4.2072\ k_BT \frac{h}{h_0^2} \quad (4.36b)$$

In order to compare models A and B, consider the case where the initial distance in model B is chosen so that $h_0 = b\sqrt{N}$, where N is the chain lengths in model A. By performing this transformation, in the short distance limit, Eq. 4.30b becomes the same as Eq. 4.12b, obtained for model A. This shows that both models have the same behavior at small distances.

In the large distance limit both models have linear force-distance relationship, but the slope in model B is larger than that in model A. This means that assuming the condition $h_0 = b\sqrt{N}$, polydisperse model is stronger than monodisperse at large distances.

4.4. Conclusion

In this chapter, we have studied polymer mediated interaction between two parallel plates. The problem involves a thin polymer layer that bridges two plates. We considered two models. In one, all polymers have the same contour length. In the other, the distribution of the bridge contour lengths is governed by the unbiased statistics of the chains at a specific distance h_0 between the plates.

We modeled the polymers as ideal chains and ignored the excluded volume effect. We have also assumed a dilute regime such that inter-chain interactions are neglected. We then have derived an analytical expression for the force that is required to hold the system in equilibrium, when the distance between two plates is slowly changed.

For both models we found repulsion at small distances and attraction at large distances, where by small and large we mean compared to $b\sqrt{N}$ for model A, and to the initial distance h_0 for model B. The force in the short distance limit is inversely proportional to the distance cubed. The proportionality constant is the same for both models as well as for the cases of tail and loops. This means that the details of the polymer layer are not important at short distances.

At large distances, the force between two plates in both models is attractive and linearly proportional to distance. The proportionality constant is not universal but is model-dependent.

Here, we only considered bridges between two plates. In general, a combination of bridges and loops can exist between two plates. The derivation presented in this chapter can be straightforwardly extended to those cases.

Chapter Five

Closure

Reduced-dimensionality, coarse-grained models are commonly employed to describe the structure and dynamics of large molecular systems. These models can bridge the gap in the time and length scale between the atomistic simulation and real biological processes. In this dissertation, we studied three different types of such models.

In the second chapter, we presented a rigorous coarse-graining method for dynamics of linear systems. The method relies on dividing the configurational space of the original atomistic model into masters (the degrees of freedom to be retained in the coarse-grained model) and slaves (the degrees of freedom to be eliminated). The method results in hierarchy of approximations and gives rise to effective stiffness, friction, mass and higher order properties for the system. The method, further, reveals that the correct effective friction for the coarse-grained model is non-diagonal, meaning that coarse-graining results in hydrodynamic-like coupling between the coarse-grained degrees of freedom. We showed by analyzing the slow relaxation modes of proteins that this dynamical coupling, which is often neglected in *ad hoc* coarse-graining procedures, is crucial in capturing relaxation modes and relaxation rates of the proteins.

The accuracy of the coarse-graining method presented in chapter 2 depends on how the master and slave degrees of freedom are chosen. The best result is obtained when the masters are slow and slaves are fast degrees of freedom. In this regard, an interesting direction for future research is to identify optimal partitioning into masters and slaves.

Also it is interesting to investigate extensions of the method to nonlinear systems. In fact, the proposed coarse-graining method can be considered as a reasonable start for coarse-graining non-linear systems. To this end, one may exploit the effective non-diagonal friction matrix obtained by linear approximation, as described in Chapter 2, along with the nonlinear potential of mean force estimated using standard coarse-graining techniques.

In Chapter 3, we evaluated the calibration procedure in elastic network models. These models have been widely used to describe the low-frequency collective motions in proteins. Elastic network models are often calibrated by fitting the mean-square displacements obtained from X-ray crystallography experiments (B-factors). However, the rigid body motion of the protein in the crystalline environment is often neglected in the calibration procedure. We designed an elastic network model which takes both the deformation and rigid body motion into account. Fitting the new model to the crystallographic B-factors revealed that B-factors are dominated by the rigid body motion of the protein. We further showed, by studying a data set of 330 proteins, that B-factors could well be described by a pure geometrical model, the translation libration screw (TLS) model, which treats proteins as rigid bodies. We also proposed an extended TLS models that treats the bulk of protein as rigid body, but allows for the flexibility of the ends. This model outperforms other models in describing B-factors. We emphasize that in this study, we did not question the usefulness of the elastic network models in any way. Rather, we expose the flaws in the current calibration procedure based on B-factors. A

better approach for the calibration of elastic networks could be based upon molecular dynamics.

In Chapter 4, we studied polymer mediated interaction between two parallel plates. In particular, we considered the case where a thin polymer layer bridges two plates. Understanding this problem is crucial in studying cell adhesion, colloid stability, and many more phenomena in biophysics and polymer science. We considered polymers with monodisperse and polydisperse chains. For both cases, by modeling polymer molecules as ideal chains, we derived analytical expressions for the force-distance relationship for the two plates. The results showed that in both cases, the plates repel from each other at small thicknesses and attract each other at large thicknesses. Further, we established that the repulsion force is inversely proportional to the thickness cubed, and the attraction force is proportional to the thickness. Also we established that polydispersity is important for large thicknesses only, as it results in larger attraction forces. The study presented here is the first step in studying these two models comprehensively. For example, the accuracy of the model can be further increased by considering the excluded volume effect in modeling chains. One can also consider loops, i. e. the chains that attach at both ends to the same surface, which is formed in the second models. Loops always cause repulsive interaction between the plates and incorporating them in our second model, will probably change the behavior of the system.

Bibliography

1. Jardetzky, O. and J.-F. Lefèvre, *Protein dynamics, function, and design* 1998, New York: Plenum Press.
2. Douzou, P. and G.H.B. Hoa, *Probing enzyme dynamics and regulation by cryo-studies*, in *Schowen, R. L. And A. Barth*. 1987. p. 223-230.
3. Allen, M.P. and D.J. Tildesley, *Computer simulation of liquids*. 1987, New York: Oxford University Press.
4. Maragakis, P., et al., *Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins.* Journal of Physical Chemistry B, 2008. **112**(19): p. 6155-6158.
5. Freddolino, P.L., et al., *Ten-microsecond molecular dynamics simulation of a fast-folding WW domain.* Biophysical Journal, 2008. **94**(10): p. L75-L77.
6. Kessel, A. and N. Ben-Tal, *Introduction to Proteins : Structure, Function, and Motion*. 2011, Boca Raton, FL: CRC Press.
7. Atilgan, A.R., et al., *Anisotropy of fluctuation dynamics of proteins with an elastic network model.* Biophysical Journal, 2001. **80**(1): p. 505-515.
8. Zheng, W.J., *A unification of the elastic network model and the Gaussian network model for optimal description of protein conformational motions and fluctuations.* Biophysical Journal, 2008. **94**(10): p. 3853-3857.
9. Hills, R.D., L.Y. Lu, and G.A. Voth, *Multiscale Coarse-Graining of the Protein Energy Landscape.* Plos Computational Biology, 2010. **6**(6).
10. Rizk, S.S., et al., *Allosteric control of ligand-binding affinity using engineered conformation-specific effector proteins.* Nature Structural & Molecular Biology, 2011. **18**(4): p. 437-U69.
11. Zhang, J., et al., *Protein Folding Simulations: From Coarse-Grained Model to All-Atom Model.* Iubmb Life, 2009. **61**(6): p. 627-643.
12. Anfinsen, C.B., *Principles that govern folding of protein chains.* Science, 1973. **181**(4096): p. 223-230.
13. Gething, M.J. and J. Sambrook, *Protein folding in the cell.* Nature, 1992. **355**(6355): p. 33-45.
14. Buchner, G.S., et al., *Dynamics of protein folding: Probing the kinetic network of folding-unfolding transitions with experiment and theory.* Biochimica Et Biophysica Acta-Proteins and Proteomics. **1814**(8): p. 1001-1020.
15. Tsai, C.J. and R. Nussinov, *The building block folding model and the kinetics of protein folding.* Protein Engineering, 2001. **14**(10): p. 723-733.
16. Bowman, G.R., V.A. Voelz, and V.S. Pande, *Atomistic Folding Simulations of the Five-Helix Bundle Protein lambda(6-85).* Journal of the American Chemical Society. **133**(4): p. 664-667.
17. Freddolino, P.L., et al., *Challenges in protein folding simulations.* Nature Physics, 2010. **6**: p. 751-758.

18. Blaszczyk, J., et al., *Crystal structure of unligated guanylate kinase from yeast reveals GMP-induced conformational changes.* Journal of Molecular Biology, 2001. **307**(1): p. 247-257.
19. Derreumaux, P. and T. Schlick, *The loop opening/closing motion of the enzyme triosephosphate isomerase.* Biophysical Journal, 1998. **74**(1): p. 72-81.
20. Huang, K.C., C. Vega, and A. Gopinathan, *Conformational changes, diffusion and collective behavior in monomeric kinesin-based motility.* Journal of Physics-Condensed Matter, 2011. **23**(37).
21. Alder, B.J. and T.E. Wainwright, *Phase Transition for a Hard Sphere System.* Journal of Chemical Physics, 1957. **27**(5): p. 1208.
22. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins.* Nature, 1977. **267**(5612): p. 585-590.
23. MacKerell, A.D., et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins.* Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
24. Jorgensen, W.L. and J. Tiradorives, *The OPLS potential functions for proteins - Energy minimizations for crystals of cyclic-peptides and Crambin.* Journal of the American Chemical Society, 1988. **110**(6): p. 1657-1666.
25. Weiner, P.K. and P.A. Kollman, *Amber-assisted model-building with energy refinement - a general program for modeling molecules and their interactions.* Journal of Computational Chemistry, 1981. **2**(3): p. 287-303.
26. Zwanzig, R., *Nonequilibrium Statistical Mechanics*. 2001, New York: Oxford Univ. Press.
27. Camilloni, C., R.A. Broglia, and G. Tiana, *Hierarchy of folding and unfolding events of protein G, CI2, and ACBP from explicit-solvent simulations.* Journal of Chemical Physics. **134**(4).
28. Higo, J., Y. Nishimura, and H. Nakamura, *A Free-Energy Landscape for Coupled Folding and Binding of an Intrinsically Disordered Protein in Explicit Solvent from Detailed All-Atom Computations.* Journal of the American Chemical Society. **133**(27): p. 10448-10458.
29. Karino, Y. and N. Matubayasi, *Free-energy analysis of hydration effect on protein with explicit solvent: Equilibrium fluctuation of cytochrome c.* Journal of Chemical Physics, 2011. **134**(4).
30. Levy, R.M. and E. Gallicchio, *Computer simulations with explicit solvent: Recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects.* Annual Review of Physical Chemistry, 1998. **49**: p. 531-567.
31. Feig, M. and C.L. Brooks, *Recent advances in the development and application of implicit solvent models in biomolecule simulations.* Current Opinion in Structural Biology, 2004. **14**(2): p. 217-224.
32. Roux, B. and T. Simonson, *Implicit solvent models.* Biophysical Chemistry, 1999. **78**(1-2): p. 1-20.
33. Bashford, D. and D.A. Case, *Generalized born models of macromolecular solvation effects.* Annual Review of Physical Chemistry, 2000. **51**: p. 129-152.

34. Onufriev, A., D. Bashford, and D.A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized born model.* Proteins-Structure Function and Bioinformatics, 2004. **55**(2): p. 383-394.
35. Onufriev, A., D. Bashford, and D.A. Case, *Modification of the generalized Born model suitable for macromolecules.* Journal of Physical Chemistry B, 2000. **104**(15): p. 3712-3720.
36. Paul, W., D.Y. Yoon, and G.D. Smith, *An optimized united atom model for simulations of polymethylene melts.* Journal of Chemical Physics, 1995. **103**(4): p. 1702-1709.
37. Smith, G.D. and W. Paul, *United atom force field for molecular dynamics simulations of 1,4-polybutadine based on quantum chemistry calculations on model molecules.* Journal of Physical Chemistry A, 1998. **102**(7): p. 1200-1208.
38. Ahmed, A. and H. Gohlke, *Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory.* Proteins-Structure Function and Bioinformatics, 2006. **63**(4): p. 1038-1051.
39. Jang, H., S. Na, and K. Eom, *Multiscale network model for large protein dynamics.* Journal of Chemical Physics, 2009. **131**(24): p. 245106.
40. Bahar, I. and A.J. Rader, *Coarse-grained normal mode analysis in structural biology.* Current Opinion in Structural Biology, 2005. **15**(5): p. 586-592.
41. Miller, B.T., et al., *Langevin network model of myosin.* Journal of Physical Chemistry B, 2008. **112**(19): p. 6274-6281.
42. Voth, G.A., *Coarse-Graining of Condensed Phase and Biomolecular Systems*. 2009, Boca Raton: CRC Press.
43. Marrink, S.J., et al., *The MARTINI Force Field.* Coarse-Graining of Condensed Phase and Biomolecular Systems, 2009: p. 5-19.
44. Zhou, J., et al., *Coarse-grained peptide modeling using a systematic multiscale approach.* Biophysical Journal, 2007. **92**(12): p. 4289-4303.
45. Noid, W.G., et al., *The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models.* Journal of Chemical Physics, 2008. **128**(24).
46. Noid, W.G., et al., *The Multiscale Coarse-Graining Method: A Systematic Approach to Coarse-Graining.* Coarse-Graining of Condensed Phase and Biomolecular Systems, 2009: p. 21-39.
47. Liu, P., S. Izvekov, and G.A. Voth, *Multiscale coarse-graining of monosaccharides.* Journal of Physical Chemistry B, 2007. **111**: p. 11566-11575.
48. Izvekov, S. and G.A. Voth, *A multiscale coarse-graining method for biomolecular systems.* Journal of Physical Chemistry B, 2005. **109**(7): p. 2469-2473.
49. Izvekov, S., P.W. Chung, and B.M. Rice, *The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials.* Journal of Chemical Physics, 2010. **133**(6).
50. Cieplak, M. and T.X. Hoang, *Scaling of folding properties in Go models of proteins.* Journal of Biological Physics, 2000. **26**(4): p. 273-294.

51. Hills, R.D., Jr. and C.L. Brooks, III, *Insights from Coarse-Grained Go Models for Protein Folding and Dynamics.* International Journal of Molecular Sciences, 2009. **10**(3): p. 889-905.
52. Paci, E., M. Vendruscolo, and M. Karplus, *Validity of Go models: Comparison with a solvent-shielded empirical energy decomposition.* Biophysical Journal, 2002. **83**(6): p. 3032-3038.
53. Chennubhotla, C., et al., *Elastic network models for understanding biomolecular machinery: From enzymes to supramolecular assemblies.* Physical Biology, 2005. **2**(4): p. S173-S180.
54. Dietz, H. and M. Rief, *Elastic bond network model for protein unfolding mechanics.* Physical Review Letters, 2008. **100**(9).
55. Hastings, W.A., et al., *Enhanced elastic network model for tRNA folding*, in *51st Annual Meeting of the Biophysical-Society*2007: Baltimore, MD. p. 227A-227A.
56. Jernigan, R.L., et al., *Elastic Network Models of Coarse-Grained Proteins Are Effective for Studying the Structural Control Exerted over Their Dynamics.* Coarse-Graining of Condensed Phase and Biomolecular Systems, 2009: p. 237-254.
57. Ryan, P., et al., *Elastic Network Model of a Nuclear Transport Complex.* Iscm Ii and Epmesc Xii, Pts 1 and 2, 2010. **1233**: p. 791-794.
58. Bahar, I., et al., *Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins.* Chemical Reviews, 2010. **110**(3): p. 1463-1497.
59. Fuchigami, S., et al., *Normal mode analysis of protein dynamics in a non-Eckart frame.* Journal of Chemical Physics, 2010. **132**(10).
60. Kaledin, M., et al., *Normal mode analysis using the driven molecular dynamics method. II. An application to biological macromolecules.* Journal of Chemical Physics, 2004. **121**(12): p. 5646-5653.
61. Maisuradze, G.G., X. Yu, and D.M. Leitner, *Normal mode analysis and calculation of the cooling rates of the chromophore vibrations during isomerization of photoactive yellow protein.* Journal of Biological Physics and Chemistry, 2007. **7**(1): p. 25-29.
62. Phillips, G.N., Jr., *Normal Mode Analysis in Studying Protein Motions with X-Ray Crystallography.* Chapman & Hall/CRC Mathematical and Computational Biology Series, 2006: p. 155-169.
63. Brooks, B.R., D. Janezic, and M. Karplus, *Harmonic-analysis of large systems .1. Methodology.* Journal of Computational Chemistry, 1995. **16**(12): p. 1522-1542.
64. Miyashita, O. and F. Tama, *Coarse-Grained Normal Mode Analysis to Explore Large-Scale Dynamics of Biological Molecules.* Coarse-Graining of Condensed Phase and Biomolecular Systems, 2009: p. 267-284.
65. Delarue, M. and Y.H. Sanejouand, *Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the Elastic Network Model.* Journal of Molecular Biology, 2002. **320**(5): p. 1011-1024.
66. Tama, F. and Y.H. Sanejouand, *Conformational change of proteins arising from normal mode calculations.* Protein Engineering, 2001. **14**(1): p. 1-6.

67. Ma, J.P., *Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes.* Structure, 2005. **13**(3): p. 373-380.
68. Essiz, S.G. and R.D. Coalson, *Langevin dynamics of molecules with internal rigid fragments in the harmonic regime.* Journal of Chemical Physics, 2007. **127**.
69. Lamm, G. and A. Szabo, *Langevin modes of macromolecules.* Journal of Chemical Physics, 1986. **85**(12): p. 7334-7348.
70. Zhou, L. and S.A. Siegelbaum, *Effects of surface water on protein dynamics studied by a novel coarse-grained normal mode approach.* Biophysical Journal, 2008. **94**(9): p. 3461-3474.
71. Caballero-Manrique, E., et al., *A theory of protein dynamics to predict NMR relaxation.* Biophysical Journal, 2007. **93**(12): p. 4128-4140.
72. Tirion, M.M., *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.* Physical Review Letters, 1996. **77**(9): p. 1905-1908.
73. Haliloglu, T., I. Bahar, and B. Erman, *Gaussian dynamics of folded proteins.* Physical Review Letters, 1997. **79**(16): p. 3090-3093.
74. Hu, M.W., et al., *Precise spring constant assignment in elastic network model for identification of vibration frequency and modeshape.* Journal of Mechanical Science and Technology, 2010. **24**(9): p. 1771-1780.
75. Jang, Y. and X.-F. Wan, *Development of Chemical Bond based Elastic Network Model and its application in identifying functional motions in H5N1 highly pathogenic avian influenza viruses.* Int J Bioinform Res Appl, 2010. **6**(1): p. 1-11.
76. Jeong, J.I. and M.K. Kim, *A chemical bond based elastic network model in proteins.* Biophysical Journal, 2005. **88**(1): p. 511A-511A.
77. Song, G. and R.L. Jernigan, *An enhanced elastic network model to represent the motions of domain-swapped proteins.* Proteins-Structure Function and Bioinformatics, 2006. **63**(1): p. 197-209.
78. Kurkcuoglu, O., R.L. Jernigan, and P. Doruker, *Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions.* Polymer, 2004. **45**(2): p. 649-657.
79. Jang, Y.H. and M.K. Kim, *Hybrid elastic network model reveals global and local motions of ribosome*, in *51st Annual Meeting of the Biophysical-Society*2007: Baltimore, MD. p. 540A-540A.
80. Eom, K., et al., *Coarse-Grained Elastic Models of Protein Structures for Understanding Their Mechanics and Dynamics.* Journal of Computational and Theoretical Nanoscience, 2010. **7**(7): p. 1210-1226.
81. Tekpinar, M. and W. Zheng, *Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model.* Proteins-Structure Function and Bioinformatics. **78**(11): p. 2469-2481.
82. Yang, L., G. Song, and R.L. Jernigan, *Protein elastic network models and the ranges of cooperativity.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(30): p. 12347-12352.
83. Zheng, W., *A unification of the elastic network model and the Gaussian network model for optimal description of protein conformational motions and fluctuations.* Biophysical Journal, 2008. **94**(10): p. 3853-3857.

84. Williams, G., *Elastic network model of allosteric regulation in protein kinase PDK1*. BMC Structural Biology, 2010. **10**: p. Article No.: 11.
85. Lu, M.Y. and J.P. Ma, *A minimalist network model for coarse-grained normal mode analysis and its application to biomolecular x-ray crystallography.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(40): p. 15358-15363.
86. Ma, J., *Coarse-Grained Elastic Normal Mode Analysis and Its Applications in X-Ray Crystallographic Refinement at Moderate Resolutions.* Coarse-Graining of Condensed Phase and Biomolecular Systems, 2009: p. 255-266.
87. Grosberg, A.Y. and A.R. Khokhlov, *Statistical physics of macromolecules*. 1994, New York: AIP Press.
88. Soheilifard, R., D.E. Makarov, and G.J. Rodin, *Rigorous coarse-graining for the dynamics of linear systems with applications to relaxation dynamics in proteins.* Journal of Chemical Physics. **135**(5).
89. Soheilifard, R., D.E. Makarov, and G.J. Rodin, *Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors.* Phys Biol, 2008. **5**(2): p. 026008.
90. Tozzini, V., *Coarse-grained models for proteins.* Current Opinion in Structural Biology, 2005. **15**(2): p. 144-150.
91. Chaimovich, A. and M.S. Shell, *Relative entropy as a universal metric for multiscale errors.* Physical Review E, 2010. **81**(6): p. 060104.
92. Zwanzig, R., *Ensemble Method in the Theory of Irreversibility.* Journal of Chemical Physics, 1960. **33**(5): p. 1338-1341.
93. Zwanzig, R., *Memory Effects in Irreversible Thermodynamics.* Physical Review, 1961. **124**(4): p. 983-992.
94. Mori, H., *Transport Collective Motion and Brownian Motion.* Progress of Theoretical Physics, 1965. **33**(3): p. 423-&.
95. Perico, A., et al., *POSITIONAL TIME CORRELATION-FUNCTION FOR ONE-DIMENSIONAL SYSTEMS WITH BARRIER CROSSING - MEMORY FUNCTION CORRECTIONS TO THE OPTIMIZED ROUSE-ZIMM APPROXIMATION.* Journal of Chemical Physics, 1993. **98**(1): p. 564-573.
96. Chang, X.Y. and K.F. Freed, *Theory for long time polymer and protein dynamics: Tests for all-atom models of alkane dynamics.* Journal of Chemical Physics, 1996. **104**(8): p. 3092-3110.
97. Izvekov, S. and G.A. Voth, *Modeling real dynamics in the coarse-grained representation of condensed phase systems.* Journal of Chemical Physics, 2006. **125**(15): p. 151101.
98. Hayward, S., A. Kitao, and N. Go, *HARMONICITY AND ANHARMONICITY IN PROTEIN DYNAMICS - A NORMAL-MODE ANALYSIS AND PRINCIPAL COMPONENT ANALYSIS.* Proteins-Structure Function and Genetics, 1995. **23**(2): p. 177-186.
99. Hinsen, K., et al., *Harmonicity in slow protein dynamics.* Chemical Physics, 2000. **261**(1-2): p. 25-37.

100. Brooks, B. and M. Karplus, *HARMONIC DYNAMICS OF PROTEINS - NORMAL-MODES AND FLUCTUATIONS IN BOVINE PANCREATIC TRYPSIN-INHIBITOR.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1983. **80**(21): p. 6571-6575.
101. Levitt, M., C. Sander, and P.S. Stern, *PROTEIN NORMAL-MODE DYNAMICS - TRYPSIN-INHIBITOR, CRAMBIN, RIBONUCLEASE AND LYSOZYME.* Journal of Molecular Biology, 1985. **181**(3): p. 423-447.
102. Case, D.A., *NORMAL-MODE ANALYSIS OF PROTEIN DYNAMICS.* Current Opinion in Structural Biology, 1994. **4**(2): p. 285-290.
103. Suhre, K. and Y.H. Sanejouand, *ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement.* Nucleic Acids Research, 2004. **32**: p. W610-W614.
104. Kidera, A. and N. Go, *REFINEMENT OF PROTEIN DYNAMIC STRUCTURE - NORMAL MODE REFINEMENT.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(10): p. 3718-3722.
105. Kidera, A. and N. Go, *NORMAL MODE REFINEMENT - CRYSTALLOGRAPHIC REFINEMENT OF PROTEIN DYNAMIC STRUCTURE .1. THEORY AND TEST BY SIMULATED DIFFRACTION DATA.* Journal of Molecular Biology, 1992. **225**(2): p. 457-475.
106. Kidera, A., et al., *NORMAL MODE REFINEMENT - CRYSTALLOGRAPHIC REFINEMENT OF PROTEIN DYNAMIC STRUCTURE .2. APPLICATION TO HUMAN LYSOZYME.* Journal of Molecular Biology, 1992. **225**(2): p. 477-486.
107. Smith, J., et al., *INELASTIC NEUTRON-SCATTERING ANALYSIS OF LOW-FREQUENCY MOTIONS IN PROTEINS - HARMONIC AND DAMPED HARMONIC MODELS OF BOVINE PANCREATIC TRYPSIN-INHIBITOR.* Journal of Chemical Physics, 1990. **93**(5): p. 2974-2991.
108. Kneller, G.R., *Inelastic neutron scattering from damped collective vibrations of macromolecules.* Chemical Physics, 2000. **261**(1-2): p. 1-24.
109. Togashi, Y. and A.S. Mikhailov, *Nonlinear relaxation dynamics in elastic networks and design principles of molecular machines.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(21): p. 8697-8702.
110. Togashi, Y., T. Yanagida, and A.S. Mikhailov, *Nonlinearity of Mechanochemical Motions in Motor Proteins.* Plos Computational Biology, 2010. **6**(6): p. e1000814.
111. Flechsig, H. and A.S. Mikhailov, *Tracing entire operation cycles of molecular motor hepatitis C virus helicase in structurally resolved dynamical simulations.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(49): p. 20875-20880.
112. Doruker, P., R.L. Jernigan, and I. Bahar, *Dynamics of large proteins through hierarchical levels of coarse-grained structures.* Journal of Computational Chemistry, 2002. **23**(1): p. 119-127.
113. Kurkcuoglu, O., R.L. Jernigan, and P. Doruker, *Collective dynamics of large proteins from mixed coarse-grained elastic network model.* Qsar & Combinatorial Science, 2005. **24**(4): p. 443-448.

114. Lu, M.Y., B. Poon, and J.P. Ma, *A new method for coarse-grained elastic normal-mode analysis.* Journal of Chemical Theory and Computation, 2006. **2**(3): p. 464-471.
115. Zheng, W.J., B.R. Brooks, and G. Hummer, *Protein conformational transitions explored by mixed elastic network models.* Proteins-Structure Function and Bioinformatics, 2007. **69**(1): p. 43-57.
116. Thorpe, M.F., *Comment on elastic network models and proteins.* Physical Biology, 2007. **4**(1): p. 60-63.
117. Bahar, I., C. Chennubhotla, and B. Erman, *Reply to 'Comment on elastic network models and proteins'.* Physical Biology, 2007. **4**(1): p. 64-65.
118. Chennubhotla, C. and I. Bahar, *Markov methods for hierarchical coarse-graining of large protein dynamics.* Journal of Computational Biology, 2007. **14**(6): p. 765-776.
119. Soheilifard, R., D.E. Makarov, and G.J. Rodin, *Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors.* Physical Biology, 2008. **5**(2): p. 026008.
120. Dietz, H. and M. Rief, *Elastic bond network model for protein unfolding mechanics.* Physical Review Letters, 2008. **100**(9): p. 098101.
121. Stember, J.N. and W. Wriggers, *Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion.* Journal of Chemical Physics, 2009. **131**(7): p. 074112.
122. Hicks, S.D. and C.L. Henley, *Coarse-grained protein-protein stiffnesses and dynamics from all-atom simulations.* Physical Review E, 2010. **81**(3): p. 030903.
123. Ming, D. and M.E. Wall, *Allostery in a coarse-grained model of protein dynamics.* Physical Review Letters, 2005. **95**(19): p. 198103.
124. Melzak, K.A., et al., *Why size and speed matter: frequency dependence and the mechanical properties of biomolecules.* Soft Matter, 2011. **7**(2): p. 332-342.
125. Wang, Y. and G. Zocchi, *Elasticity of Globular Proteins Measured from the ac Susceptibility.* Physical Review Letters, 2010. **105**(23): p. 238104.
126. Faccioli, P., A. Lonardi, and H. Orland, *Dominant reaction pathways in protein folding: A direct validation against molecular dynamics simulations.* Journal of Chemical Physics, 2010. **133**(4): p. 045104.
127. Eom, K., et al., *Modified mechanical mass-spring model of biomolecules for understanding dynamics of proteins.* Journal of Mechanical Science and Technology, 2008. **22**(3): p. 506-513.
128. Eom, K., et al., *Coarse-graining of protein structures for the normal mode studies.* Journal of Computational Chemistry, 2007. **28**(8): p. 1400-1410.
129. Kim, J.I., S. Na, and K. Eom, *Domain Decomposition-Based Structural Condensation of Large Protein Structures for Understanding Their Conformational Dynamics.* Journal of Computational Chemistry, 2010. **32**(1): p. 161-169.
130. Ermak, D.L. and J.A. McCammon, *BROWNIAN DYNAMICS WITH HYDRODYNAMIC INTERACTIONS.* Journal of Chemical Physics, 1978. **69**(4): p. 1352-1360.

131. Ponder, J.W., *Tinker- Software Tools for Molecular Design*, 2003, Washington University School of Medicine: Saint Louis, MO.
132. Pastor, R.W. and M. Karplus, *PARAMETRIZATION OF THE FRICTION CONSTANT FOR STOCHASTIC SIMULATIONS OF POLYMERS.* Journal of Physical Chemistry, 1988. **92**(9): p. 2636-2641.
133. Rotne, J. and S. Prager, *VARIATIONAL TREATMENT OF HYDRODYNAMIC INTERACTION IN POLYMERS.* Journal of Chemical Physics, 1969. **50**(11): p. 4831-&.
134. De La Torre, J.G. and V.A. Bloomfield, *Hydrodynamic Properties of Macromolecular Complexes. I. Translation.* Biopolymers, 1977. **16**(8): p. 1747-1763.
135. Hills, R.D., L.Y. Lu, and G.A. Voth, *Multiscale Coarse-Graining of the Protein Energy Landscape.* Plos Computational Biology, 2010. **6**(6): p. e1000827.
136. Zhang, Z.Y., et al., *A Systematic Methodology for Defining Coarse-Grained Sites in Large Biomolecules.* Biophysical Journal, 2008. **95**(11): p. 5073-5083.
137. Guyan, R.J., *Reduction of Stiffness and Mass Matrices.* AIAA Journal, 1965. **3**(2): p. 380.
138. Elber, R., *Long-timescale simulation methods.* Current Opinion in Structural Biology, 2005. **15**: p. 151-156.
139. Nielsen, S.O., et al., *Coarse grain models and the computer simulation of soft materials.* Journal of Physics-Condensed Matter, 2004. **16**(15): p. R481-R512.
140. Brooks, B. and M. Karplus, *Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme.* Proc Natl Acad Sci U S A, 1985. **82**: p. 4995-4999.
141. Levitt, M., C. Sander, and P.S. Stern, *Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme,.* J Mol Biol, 1985. **181**: p. 423-447.
142. Ma, J., *Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes.* Structure, 2005. **13**(3): p. 373-380.
143. Ma, J. and M. Karplus, *Ligand-induced conformational changes in ras p21: a normal mode and energy minimization analysis.* J Mol Biol, 1997. **274**(1): p. 114-31.
144. Seno, Y. and N. Go, *Deoxymyoglobin studied by the conformational normal mode analysis. I. Dynamics of globin and the heme-globin interaction.* J Mol Biol, 1990. **216**: p. 95-109.
145. Andricioaei, I. and M. Karplus, *On the calculation of entropy from covariance matrices of the atomic fluctuations.* J. Chem. Phys., 2001. **115**(14): p. 6289.
146. Schafer, H., A.E. Mark, and W.F. van Gunsteren, *Absolute entropies from molecular dynamics simulation trajectories.* J. Chem. Phys., 2000. **113**(18): p. 7809.
147. Tirion, M.M., *Large Amplitude Elastic Motions in Proteins from a Single-arameter, Atomic Analysis.* Phys. Rev. Letters, 1996. **77**: p. 1905.
148. Atilgan, A.R., et al., *Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model.* Biophysical Journal, 2001. **80**: p. 505-515.

149. Bahar, I., C. Chennubhotla, and B. Erman, *Reply to 'Comment on elastic network models and proteins'.* Physical Biology, 2007. **4**: p. 64-65.
150. Bahar, I. and A.J. Rader, *Coarse grained normal modes in structural biology.* Current Opinion in Structural Biology, 2005. **15**: p. 586-592.
151. Chennubhotla, C., et al., *Elastic Network Models for Understanding Biomolecular machinery: From Enzymes to Supramolecular Assemblies.* Physical Biology, 2005. **2**: p. S173-S180.
152. Doruker, P., R.L. Jernigam, and I. Bahar, *Dynamics of Large Proteins through Hierarchical Levels of Coarse-Grained Structures.* J. of Computational Chemistry, 2002. **23**(1): p. 119-127.
153. Erman, B. and K.A. Dill, *Gaussian Model of protein folding.* J. Chem. Phys., 2000. **112**(2): p. 1050.
154. Eyal, E., L.-W. Yang, and I. Bahar, *Anisotropic network model: systematic evaluationand a new web interface.* Bioinformatics, 2006. **22**(21): p. 2619-2627.
155. Song, G. and R.L. Jernigan, *vGNM: A better model for understanding the dynamics of proteins in crystals.* J Mol Biol, 2007. **369**(3): p. 880-893.
156. Tama, F., W. Wriggers, and C.L. Brooks III, *Exploring global distortions of biological macromolecules from low-resolution structural information and elastic network theory.* J Mol Biol, 2002. **321**: p. 297-305.
157. Thorpe, M.F., *Comment on elastic network models and proteins.* Physical Biology, 2007. **4**: p. 60-63.
158. Haliloglu, T., I. Bahar, and B. Erman, *Gaussian Dynamics of Folded Proteins.* Phys. Rev. Letters, 1997. **79**: p. 3090-3093.
159. Przemieniecki, J.S., *Theory of matrix structural analysis*. 1985, New York Dover.
160. Makarov, D.E., et al., *How the folding rate constant of simple, single-domain proteins depends on the number of native contacts.* Proc. Natl. Acad. Sci. USA, 2002. **99**: p. 3535.
161. Makarov, D.E. and H. Metiu, *A model for the kinetics of protein folding: Kinetic Monte Carlo simulations and analytical results.* J. Chem. Phys., 2002. **116**: p. 5205.
162. Makarov, D.E. and K.W. Plaxco, *The topomer search model: a quantitive, fisrt principles description of two-state protein folding kinetics.* Protein Science, 2003. **12**: p. 17-26.
163. Eom, K., et al., *Relationship between the mechanical properties and topology of cross-linked polymer molecules: Parallel strands maximize the strength of model polymers and protein domains.* J. Phys. Chem. B, 2003. **107**: p. 8730.
164. Eom, K., D.E. Makarov, and G.J. Rodin, *Theoretical studies of the kinetics of mechanical unfolding of cross-linked polymer chains and their implications for single molecule pulling experiments.* Phys. Rev. E, 2005. **71**: p. 021904.
165. Makarov, D.E. and G.J. Rodin, *Configurational entropy and mechanical properties of cross-linked polymer chains: Implications for protein and RNA folding.* Phys. Rev. E, 2002. **66**: p. 011908.
166. Tatham, A.S. and P.R. Shewry, *Elastomeric proteins: biological roles, structures and mechanisms.* TIBS, 2000. **Nov. 2000**: p. 567.

167. Miyashita, O., J.N. Onuchic, and P.G. Wolynes, *Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins.* Proc. Natl. Acad. Sci USA, 2003. **100**: p. 12570-12575.
168. Ikeguchi, M., et al., *Protein structural change upon ligand binding: linear response theory.* Phys. Rev. Letters, 2005. **94**: p. 078102.
169. Demirel, M.C. and A.M. Lesk, *Molecular forces in antibody maturation.* Phys. Rev. Letters, 2005. **95**: p. 208106.
170. Gohlke, H. and M.F. Thorpe, *A natural coarse graining for simulating large biomolecular motion.* Biophys. J., 2006. **91**: p. 2115-2120.
171. Zheng, W., B.R. Brooks, and G. Hummer, *Protein conformational transitions explored by mixed elastic network models* PROTEINS: Structure, Function and Bioinformatics, 2007. **69**(1): p. 43-57.
172. Erman, B., *The gaussian network model: precise prediction of residue fluctuations and application to binding problems.* Biophys J, 2006. **91**(10): p. 3589-99.
173. Zheng, W., *A unification of the elastic network model and the gaussian network model for optimal description of protein conformational motions and fluctuations.* Biophys J, 2008. **94**: p. 3853-3857.
174. Dietz, H. and M. Rief, *Elastic bond network model for protein unfolding mechanics.* Phys Rev Lett, 2008. **100**(9): p. 098101.
175. Eyal, E. and I. Bahar, *Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models.* Biophys J, 2008. **94**(9): p. 3424-35.
176. Ming, D. and M. Wall, *Allostery in a coarse-grained model of protein dynamics.* Phys. Rev. Letters, 2005. **95**: p. 198103.
177. Moritsugu, K. and J.C. Smith, *Coarse-Grained Biomolecular Simulation with REACH: Realistic Extension Algorithm via Covariance Hessian.* Biophys. J., 2007. **93**: p. 3460-3469.
178. Ernan, E., L.-W. Yang, and I. Bahar, *Anisotropic network model: systematic evaluation and a new web interface.* Bioinformatics, 2006. **22**: p. 2619-2627.
179. Halle, B., *Flexibility and packing in proteins.* Proc Natl Acad Sci U S A, 2002. **99**: p. 1274-1279.
180. Schomaker, V. and K.N. Trueblood, *On the rigid-body motions of molecules in crystals.* Acta Crystallographica B, 1968. **24**: p. 63-76.
181. Sternberg, M.J.E., D.E.P. Grace, and D.C. Phillips, *Dynamic information from protein crystallography: an analysis of temperature factors from refinement of the hen egg white lysozyme structure.* J Mol Biol, 1979. **130**: p. 231-253.
182. Kuriyan, J. and W.I. Weis, *Rigid protein motion as a model for crystallographic temperature factors.* Proc Natl Acad Sci U S A, 1991. **88**: p. 2773-2777.
183. Kundu, S., et al., *Dynamics of proteins in crystals: comparison of experiments with simple models.* Biophys. J., 2002. **83**: p. 723-731.
184. Chong, S.-H., et al., *Dynamical transition of myoglobin in a crystal: comparative studies of X-ray crystallography and Mossbauer spectroscopy* Eur Biophys J, 2001. **30**: p. 319-329.

185. Kidera, A. and N. Go, *Normal mode refinement: Crystallographic refinement of protein structure. I. Theory and test by simulated diffraction data.* J Mol Biol, 1992. **225**: p. 457-475.
186. Kidera, A., et al., *Normal mode refinement: crystallographic refinement of protein dynamics structure, II. Application to human lysozyme.* J Mol Biol, 1992. **225**: p. 477-486.
187. Rader, A.J., et al., *Protein unfolding: Rigidity lost.* Proc. Natl. Acad. Sci USA, 2002. **99**(6): p. 3540-3545.
188. Phillips, G.N., *Comparison of the dynamics of myoglobin in different crystal forms.* Biophys. J., 1990. **57**: p. 381-383.
189. Zeng, J., H.R. Treutlein, and T. Simonson, *Molecular dynamics simulations of the Ras:Raf and Rap:Raf complexes.* PROTEINS: Structure, Function, and Genetics, 1999. **35**: p. 89-100.
190. Painter, J. and E.A. Merritt, *Optimal description of a protein structure in terms of multuiple groups undergoing TLS motion.* Acta Crystallographica D, 2006. **62**: p. 439-450.
191. Burden, C.J. and A.J. Oakley, *Anisotropic atomic motions in high-resolution protein crystallography molecular dynamics simulations.* Physical Biology, 2007. **4**: p. 79-90.
192. Fleer, G.J., et al., *Polymers at interfaces*. 1993, New York: Chapman & Hall.
193. Jones, R.A.L., *Polymers at surfaces and interfaces*. 1999, New York: Cambridge University Press.
194. Allegra, G., G. Raos, and C. Manassero, *Polymer-mediated adhesion: A statistical approach.* Journal of Chemical Physics, 2003. **119**(17): p. 9295-9307.
195. Barrett, S.P., *Protein mediated adhesion of staphyloccus-aureus to silicone implant polymer.* Journal of Medical Microbiology, 1985. **20**(2): p. 249-253.
196. Helgeson, M.E. and N.J. Wagner, *Colloidal interactions mediated by end-adsorbing polymer-like micelles.* THE JOURNAL OF CHEMICAL PHYSICS. **135**(8): p. 084901.
197. Joanny, J.F., L. Leibler, and P.G. Degennes, *Effects of polymers-solutions on colloid stability.* Journal of Polymer Science Part B-Polymer Physics, 1979. **17**(6): p. 1073-1084.
198. Dolan, A.K. and S.F. Edwards, *Theory of stabilization of colloids by adsorbed polymer.* Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences, 1974. **337**(1611): p. 509-516.
199. Jenkel, E. and B. Rumbach, *Adsorption of high polymers from solution.* Z. Elektrochem., 1951. **55**: p. 612.
200. Kleshchanok, D., R. Tuinier, and P.R. Lang, *Direct measurements of polymer-induced forces.* Journal of Physics-Condensed Matter, 2008. **20**(7).
201. Jiang, J.W., H.L. Liu, and Y. Hu, *Lattice Monte Carlo simulation of polymer adsorption at an interface, 2 - Polydisperse polymer.* Macromolecular Theory and Simulations, 1998. **7**(1): p. 113-117.

202. Testard, V., J. Oberdisse, and C. Ligoure, *Monte Carlo simulations of colloidal pair potential induced by telechelic polymers: Statistics of loops and bridges.* Macromolecules, 2008. **41**(19): p. 7219-7226.
203. Allegra, G., G. Raos, and C. Manassero, *Polymer chains and networks in narrow slits - Highlights of recent results*, in *Computer Simulations of Liquid Crystals and Polymers*. 2005. p. 249-268.
204. Bhatia, S.R. and W.B. Russel, *End-capped associative polymer chains between nanospheres: Attractions in ideal solutions.* Macromolecules, 2000. **33**(15): p. 5713-5720.
205. Dolan, A.K. and S.F. Edwards, *The Effect of Excluded Volume on Polymer Dispersant Action.* Proceedings of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences, 1975. **343**(1635): p. 427-442.
206. Carslaw, H.S. and J.C. Jaeger, *Conduction of heat in solids*. 1959, Oxford: Clarendon Press
207. Morse, P.M. and H. Feshbach, *Methods of theoretical physics*. 1953, New York: McGraw-Hill.
208. Zhang, B.W., D. Jasnow, and D.M. Zuckerman, *Transition-event durations in one-dimensional activated processes.* THE JOURNAL OF CHEMICAL PHYSICS, 2007. **126**(7): p. 074504.
209. Lubensky, D.K. and D.R. Nelson, *Driven polymer translocation through a narrow pore.* Biophysical Journal, 1999. **77**(4): p. 1824-1838.

Vita

Reza Soheilifard was born in Tehran, Iran. After completing his high school education at Nemuneh Bagher-ol-olom High School in Sabzevar, Iran, in 1995, he entered Sharif University of Technology in Tehran, Iran. He received his Bachelor of Engineering and Master of Engineering from Sharif University of Technology in 1999 and 2001, respectively. During the following years he worked on many industrial and research projects in Civil Engineering. He started his PhD at the department of Aerospace Engineering and Engineering Mechanics at the University of Texas at Austin in spring 2008. During his graduate program, he published three highly cited papers in the Journal of the Mechanics and Physics of Solids, Physical Biology and the Journal of Chemical Physics.

Email address: rsoheilifard@gmail.com

This dissertation was typed by the author.